Literacy, Language and Learning (L3) Initiative Bugesera Pilot Student Assessment Report
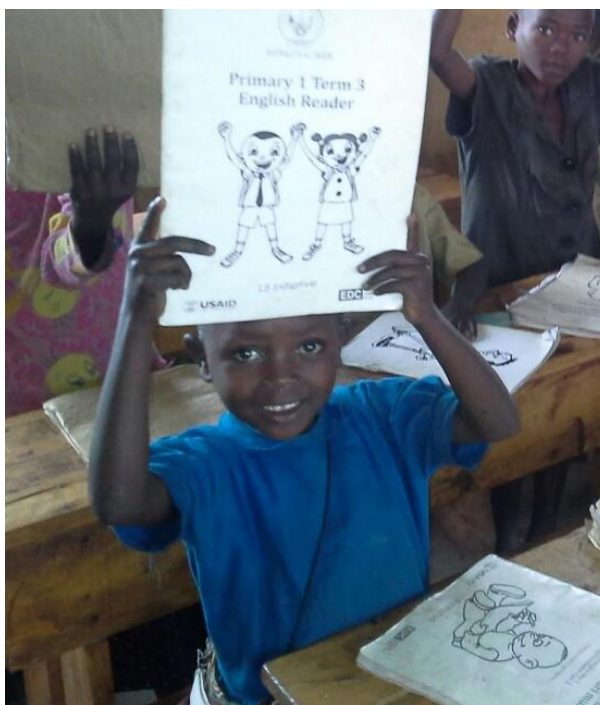
Executive Summary

The Language, Literacy and Learning Initiative (L3) is a 5-year program financed by the United States Agency for International Development (USAID) and implemented by the Education Development Center, Inc. (EDC), with technical assistance from Volunteer Services Overseas (VSO), Concern Worldwide (CW), International Educational Exchange (IEE), and Never Again Rwanda (NAR). The implementation dates are August 2011 – January 2017. The L3 initiative's goal is to strengthen teaching and learning so that children leave primary school with solid skills in literacy and math.

This report presents findings from the impact evaluation of the project's pilot literacy and numeracy intervention on students' performance in reading in Kinyarwanda and English and in math in Primary 1 (P1) and Primary 2 (P2) grades in Bugesera province. The report describes the results of the baseline and endline assessments of a sample of students in intervention schools and control schools.

The report has several purposes. One is to inform the Ministry of Education in Rwanda and others of the impact of the pilot intervention on students' performance in literacy and math. Another is to provide detailed information to those who design the literacy and math curricula and methods of teaching the specific literacy and math skills.

The study of the pilot in Bugesera found L3 intervention effective in improving student achievement in literacy and math. Data analysis showed that the intervention was particularly effective for P1 students, who scored significantly higher on the endline assessments in literacy in both languages as well as in math, compared to the control group students. P2 intervention students also scored higher than their control counterparts, though the difference between the two P2 study groups was less pronounced.

Students were tested using standardized assessments in literacy and Kinyarwanda and English, and in numeracy. The study employed cross-sectional experimental design, testing different students in the same classrooms at the baseline and the endline. Students were randomly selected from P1 and P2 grade classrooms in 12 schools (6 intervention and 6 control).The baseline assessment was conducted in March of 2012 and included 654 students. The endline assessment was conducted in September/October of 2013 and included 662 students.

Summary of Findings

Overall, the L3 pilot intervention had a notable impact on the Kinyarwanda, English and math skills of P1 and P2 students. Endline scores revealed significant differences between the L3 and control group students, with L3 students outperforming control group students.

The findings are summarized by subject: Kinyarwanda, English and math. For each subject, we first compare overall baseline and endline scores for P1 and P2 students. We then present the results of the statistical tests of significance of difference in scores between the L3 students and the control group students at the endline, and finally we summarize patterns in students' levels of proficiency.

Kinyarwanda Literacy Assessment Results

Figure 1 shows that students from L3 intervention schools demonstrated significantly larger gains in Kinyarwanda literacy between the two measurements, compared to students from the control schools. The total mean score was computed by adding up the percent correct score for each of the Kinyarwanda Literacy Assessment tasks and then dividing them by the number of tasks. The subtests of the assessment were not weighted in the composite score.

Figure 1. Overall Scores on Kinyarwanda Literacy Assessment, by Grade
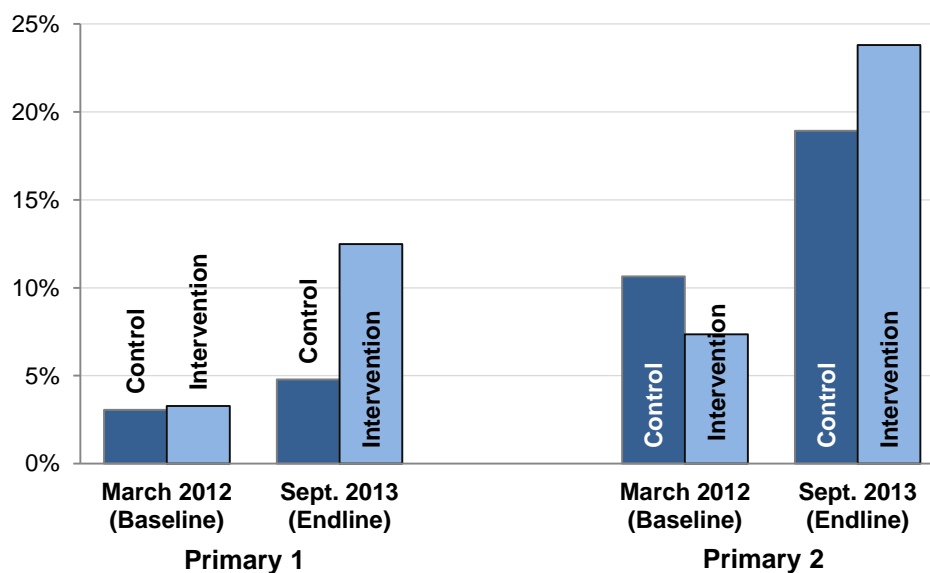
The overall scores in Figure 1 comprise scores of twelve subtests for P1 students and 12 subtests for P2 students (see list of subtests in Table 2). The comparison of means showed that the two study groups in P1 were equivalent at the baseline on all subtests, and the intervention group did better than the comparison group at the endline on all but one subtest (at $p<.001$ level). The two study groups in P2 were equivalent at baseline in three of the twelve subtests, and in the remaining nine subtests the control group students scored higher at the baseline than the intervention group students. However, at the endline the intervention group scored statistically significantly higher than the control group in seven subtests.

Table 1 presents a comparison of means in endline scores between the two groups at P1 and P2 grades. It summarizes the results of the statistical tests of significance of difference in scores (independent group *t*-test) between the L3 students and the control group students at the endline. As the table shows, the P1 intervention group (L3) students achieved statistically significantly higher scores than the control group students on all subtests but one The P2 intervention group achieved statistically significantly higher scores than the control group on seven subtests. At both grade levels, the control group scored significantly higher on the common vocabulary subtest.

Table 1. Kinyarwanda summary results of the comparison of means between L3 and control group at the endline

| | | P1 | | P2 | |
		L3	control	L3	control
Pre-Literacy skills	Common vocabulary				
	Alphabet recitation	sig. (*p*<.001)		sig. (*p*<.001)	
	Letter reading (upper case)	sig. (*p*<.001)		sig. (*p*<.05)	
	Letter reading: (lower case)	sig. (*p*<.001)		sig. (*p*<.05)	
	Letter-sound association	sig. (*p*<.001)			
	Blends	sig. (*p*<.001)		sig. (*p*<.05)	
	Counting Syllables	sig. (*p*<.001)		sig. (*p*<.001)	
	Concepts of Print	sig. (*p*<.001)			
	Nonsense word reading	sig. (*p*<.001)		sig. (*p*<.05)	
Fluency and Comprehension skills	Familiar word reading	sig. (*p*<.001)			
	Listening comprehension				
	Oral passage reading	sig. (*p*<.001)			

Baseline and endline scores on subtests. P1 students in both intervention and control study groups demonstrated moderate skill level in three out of twelve Kinyarwanda literacy assessment subtests: basic vocabulary, alphabet recitation, and counting syllables. P1 students scored over 50 percent, on average, on these three subtests at the endline. Proficiency in these three subtests ensures mastery of foundational pre-literacy skills that are necessary for reading later on. Overall, intervention students

showed statistically significantly larger gains between the pretest and the posttest in all but two subtests: basic vocabulary and listening comprehension.

Both L3 and control group P2 students scored on average over 50 percent correct on the subtests measuring pre-literacy skills, such as basic vocabulary, knowledge of the alphabet, and the concepts of print.

P2 students, on average, scored below 50 percent on all subtests aimed at measuring actual reading fluency, such as decoding unfamiliar words, reading familiar words, and reading a connected text.

The computations of fluency for both study groups showed that the control group students read on average 16.8 words per minute in a connected text passage, and intervention group students read on average 21.1 words per minute. Since fluency benchmarks for P2 have not been established yet, the study does not draw conclusions about a proportion of students meeting proficiency standards.

The examination of the assessment results revealed a high proportion of students with zero scores in individual subtests. Further data analyses showed that scores in the majority of subtests are distributed in a U-shape, with a high proportion of zero scores and/or with high scores, and with very few students achieving medium results. These results portray a diverse classroom, filled with students with different ability levels and in need of different instructional approaches. While students with zero scores might require remediation, students with high scores need access to grade level reading materials and opportunities to practice their reading skills.

English Literacy Assessment Results

Sampled P1 and P2 students were tested in English literacy skills, using an adapted Early Grade Reading Assessment. As Figure 2 illustrates, the overall level of English reading proficiency is quite low. Students from intervention schools from both P1 and P2 grades demonstrated some gains between the baseline and endline assessments, while the scores of the control group students remained flat.

Figure 2. Overall Scores on English Literacy Assessment, by Grade
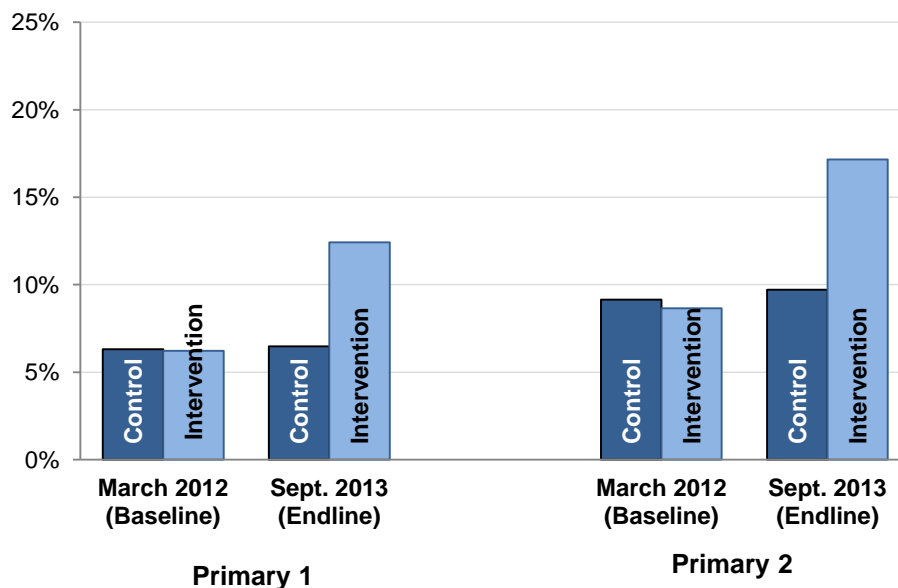
The comparison of means showed that while the two study groups in P1 were equivalent at the baseline on all but one subtest[1], the intervention group did better at the endline than the comparison group on

---

[1] Alphabet recitation, on which the control group students performed better.

all subtests. Similarly in P2, while the two study groups were equivalent at the baseline in all but two subtests of the ten subtests (basic conversation vocabulary and alphabet recitation), at the endline the intervention group scored statistically significantly higher than the control group on seven subtests.

Table 2 shows a summary of the statistical tests of significance of difference in scores (independent group *t*-test) between the L3 students and the control group students at the endline.

Table 2. English Summary results of the comparison of means between L3 and control group at the endline

		P1		P2	
		L3	**control**	**L3**	**control**
Pre-Literacy skills	Basic conversation	sig. (p<.001)		sig. (p<.001)	
	Alphabet recitation	sig. (p<.001)			
	Letter reading (upper case)	sig. (p<.001)		sig. (p<.001)	
	Letter reading: (lower case)	sig. (p<.001)		sig. (p<.001)	
	Phonemic awareness	sig. (p<.001)		sig. (p<.001)	
	Letter-sound association	sig. (p<.001)			
	Common vocabulary	sig. (p<.001)		sig. (p<.05)	
Fluency and Comprehension skills	Familiar word reading	sig. (p<.001)		sig. (p<.001)	
	Listening comprehension	sig. (p<.05)			
	Oral passage reading	sig. (p<.05)		sig. (p<.001)	

As Figure 2 shows, overall, P1 students in both intervention and control study groups demonstrated very low skill level in all English literacy assessment subtests but one: alphabet recitation. Only in the alphabet recitation subtest did any group– the intervention group – score over 50 percent correct at the endline. The next highest score was only at 30 percent correct, on the phonemic awareness subtest. Similarly, P2 students scored on average over 50 percent correct on the alphabet recitation subtest and over 20 percent only on the basic conversation skills and phonemic awareness subtests. On all subtests at both grade levels a high proportion of students had zero scores.

The results of the assessment showed a very high proportion of students with zero scores in individual subtests. Although intervention group students did show significant gains at the posttest, the overall scores remained very low.

Math Assessment Results

The data analysis of the Math assessment results showed that at the baseline control and intervention school students

performed at a similar level. However, at the endline intervention school students showed significantly better numeracy skills, compared to the control group students. Annex 4 shows details of student achievement on each subtask, by grade, including a percent of learners with zero scores. Figure 3 shows the distribution of total mean scores at the baseline and the endline for both grades. Differently from literacy figures, Figure 3 below goes up to 70 percent.

Figure 3. Overall Scores on Math Assessment, by Grade
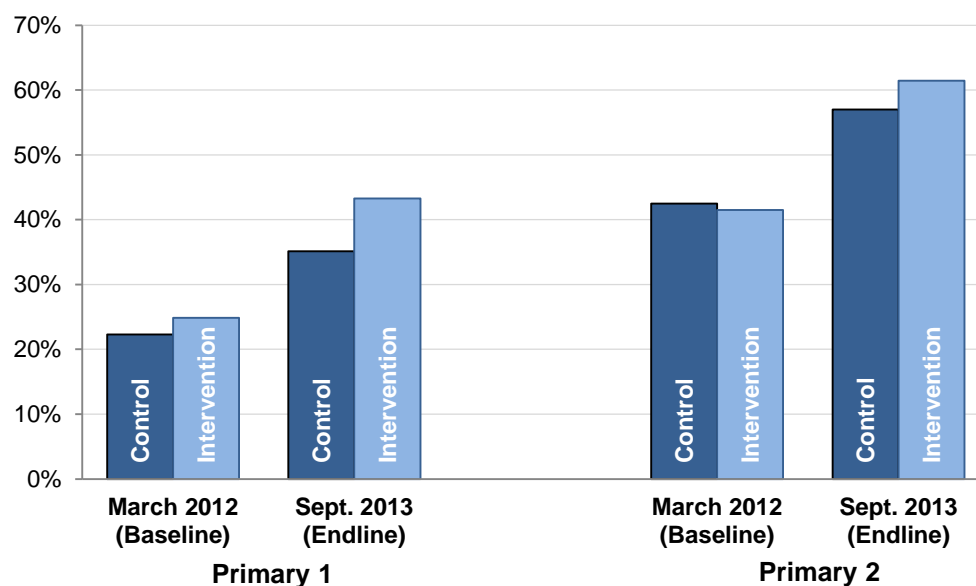
Overall, EGMA test results show that P1 and P2 students did well in most tasks on foundational math skills such as basic counting, number pairs, number identification, and basic addition and subtraction tasks using manipulatives. Students performed poorly on the following foundational math skills: counting backwards and counting by 2s and by 5s tasks. Students did not do well with tasks requiring operating numbers, such as addition and subtraction without manipulatives.

The comparison of means showed that the two study groups in P1 were not equivalent at the baseline, with the intervention group students scoring higher on five of the eleven subtests. At the endline, however, the intervention group did better than the comparison group on all subtests. P2 study groups were equivalent at the baseline, with very similar average scores on each subtest. However, at the endline the intervention group scored statistically significantly higher than the control group on four subtests. Table 3 shows a summary of the statistical tests of significance of difference in scores (independent group *t*-test) between the L3 students and the control group students at the endline.

Table 3. Summary results of the comparison of means between L3 and control group at the endline

	P1		P2	
	L3	**control**	**L3**	**control**
Counting recitation				
Counting backwards	sig. ($p<.001$)		sig. ($p<.001$)	
Counting by 2s	sig. ($p<.01$)			

Foundational math skills	Counting by 5s	sig. ($p<.001$)	sig. ($p<.05$)
	Counting by 10s	sig. ($p<.001$)	sig. ($p<.05$)
	Adding objects	sig. ($p<.001$)	
	Subtracting objects	sig. ($p<.001$)	sig. ($p<.05$)
	Number pairs	sig. ($p<.001$)	
	Number identification	sig. ($p<.05$)	
Number operations	Adding numbers	sig. ($p<.01$)	
	Subtracting numbers	sig. ($p<.01$)	

At the endline, P1 students in both intervention and control study groups scored over 50 percent, on average, in three out of eleven subtests of the assessment: adding and subtracting objects, and number pairs. While many P1 students completed individual subtests without any errors, a substantial proportion of students had zero scores. The tasks on which students demonstrated proficiency are less abstract and have relevancy in their lives. They also provide foundational skills for number operations later on, but at the time of the assessment few students demonstrated mastery of operating numbers, not objects.   Particularly noteworthy is a significant proportion of students with zero scores on the number identification subtest.

P2 students scored on average over 50 percent correct on the subtests relating to counting, adding and subtracting objects, identifying numbers and number pairs. Similar to the results of the P1 student assessment, a high proportion of students had zero scores in individual subtests. Although some subtests showed close to normal distributions of scores among students with non-zero scores, a high percent of students with zero scores skews overall results. Similar to literacy assessment results, math assessment shows that Rwandan classrooms are filled with students with very different math skills. Such diversity requires both an adjustment of instructional routine on the part of the teacher, as well as differentiated learning materials.

Summary

Overall, the impact study of the pilot in Bugesera found L3 intervention effective in improving student achievement in literacy and math. Data analysis showed that the intervention was particularly effective for P1 students who scored significantly higher on the endline assessments in literacy in both languages as well as in math, compared to the control group students. P2 intervention students also scored higher than their control counterparts, although the difference between the two P2 study groups was less pronounced.

Contents

List of Figures

List of Tables

Introduction

The Language, Literacy and Learning Initiative (L3) is a 5-year project intended to strengthen teaching and learning so that children leave primary school with solid literacy/numeracy skills. L3 is financed by the United States Agency for International Development (USAID) and implemented by the Education Development Center, Inc. (EDC), in partnership with Volunteer Services Overseas (VSO), Concern Worldwide (CW), International Educational Exchange (IEE), and Never Again Rwanda (NAR). The project began in August 2011 and is scheduled to end in January 2017. The L3 initiative's goal is to strengthen teaching and learning so that children leave primary school with solid literacy/numeracy skills.

L3 has five main objectives:

- **Improve the quality of teaching reading in Kinyarwanda and in English, and teaching mathematics in primary grades 1-4 (P1-P4).** The L3 initiative trains teachers in how to implement the new, evidence-based reading and mathematics teaching strategies and how to use associated L3 instructional materials effectively. The initiative also initiates activities to motivate teachers and improve their working conditions.

- **Improve the availability of teaching and learning materials.** The L3 initiative provides teachers with instructional materials keyed to the new instructional practices. It also increases the availability of reading material by distributing over one million supplementary books. Finally, it introduces accessible and sustainable technologies (audio, cell phones, and video) to enrich students' learning.

- **Support the teaching of English and the transition to English as the language of instruction in P4.** The L3 initiative will produce interactive audio instruction (IAI) programs for English as a second Language to develop students' and teachers' English communication skills and support their gradual transition to English as a language of instruction.

- **Strengthen Ministry capacity.** The L3 initiative embeds literacy/numeracy specialists in the central Ministry and the 13 TTCs, providing day-to-day support in literacy/numeracy and teacher training reforms.

- **Improve equity in education.** The combination of scripted teacher lessons and IAI programs ensures that all students – those in urban areas as well as those in the most remote regions - receive equal access, each day, to quality instruction. Their teachers also receive ongoing training in effective and inclusive instructional practices. The program targets additional resources and support to students in disadvantaged areas.

This report presents findings from an evaluation study of the project's pilot literacy and numeracy intervention on students' performance in reading in Kinyarwanda and English and in math in Primary 1 (P1) and Primary 2 (P2) grades in Bugesera province. The report describes the results of the baseline and endline assessments of a sample of students in intervention schools and control schools.

During the L3 pilot in Bugesera, teachers were trained by the L3 technical team on two separate occasions: in the second week of March of 2013 and in the first week of May of 2013. Both trainings were about 16 hours long (2 days) each. Prior to the first training, teaching and learning materials were distributed to 20 schools (including the 6 intervention schools). Materials included teachers' guides (for Kinyarwanda, English and Math instruction), students' daily readers (in Kinyarwanda and in English), mobile phones with memory cards that contained audio lessons in the three main subjects along with speakers to allow teachers to broadcast lessons. Students' materials were distributed on three occasions; term one in March, term two in June and term three in September. All the intervention schools also received solar panels to charge the mobile phones and speakers.

The study of the pilot in Bugesera found L3 intervention effective in improving student achievement in literacy and math. Data analysis showed that the intervention was particularly effective for P1 students, who scored significantly higher on the endline assessments in literacy in both languages as well as in math, compared to the control group students. P2 intervention students also scored higher than their control counterparts, though the difference between the two P2 study groups was less pronounced.

L3-Provided MP3 Player with Speakers for Literacy Lessons

KINYARWANDA ASSESSMENT FINDINGS

Overall Kinyarwanda Assessment Findings

Students from L3 intervention schools demonstrated significantly larger gains in Kinyarwanda literacy between the two measurements, compared to students from the control schools. The total mean score was computed by adding up the percent correct score for each of the Kinyarwanda Literacy Assessment tasks and then dividing them by the number of subtests. The subtests of the assessment were not weighted in the composite score.

The graph below shows that students from L3 intervention schools demonstrated significantly larger gains between the two measurements, compared to students from the control schools. For P1, the difference between the overall mean scores at the baseline is insignificant, but at the endline P1 intervention students show average scores significantly higher than control group students (at $p<.001$ level). For P2, the intervention students scored significantly lower at the baseline (at $p<.01$ level), but significantly higher at the endline (at $p<.01$ level).

Figure 4. Overall Scores[2] Kinyarwanda, by Grade
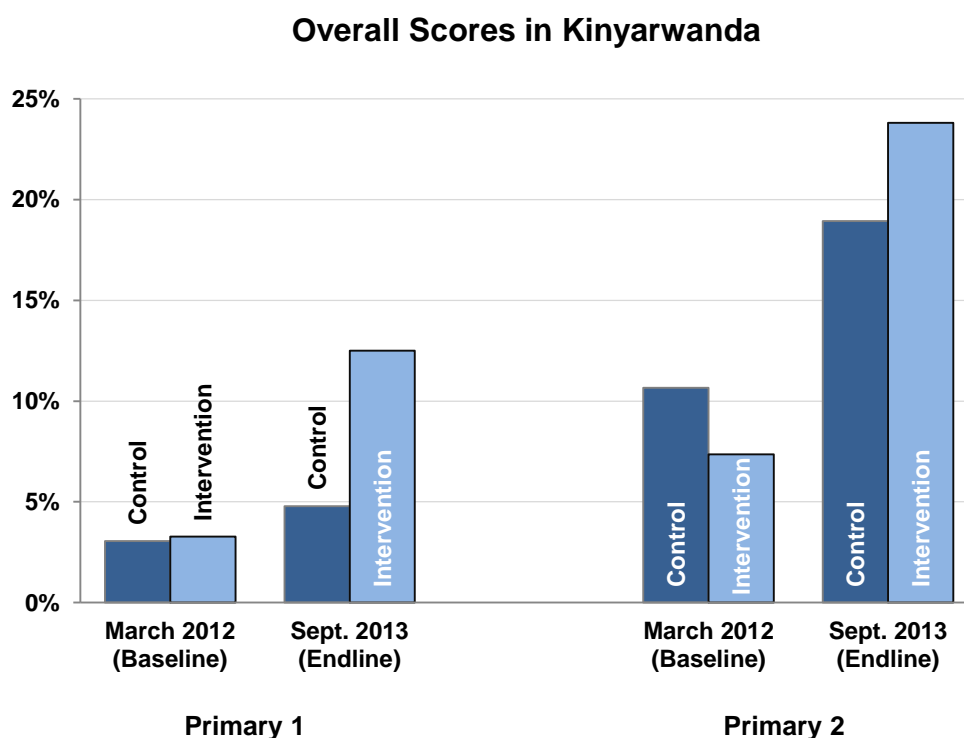
Overall Scores in Kinyarwanda

The figure below shows the distribution of the average scores at the baseline and the endline for P1 students, for each of the Kinyarwanda literacy assessment subtests. The figure shows the average

---

[2] The values in the graph are means of all subtest means.

percent correct achieved by P1 students at the baseline (blue shading) and the gain made at the endline (pink shading). The two subtests where gains are shown in stripes are those where students scored lower at the endline than at the baseline. Annex 2 shows the mean distributions for all subtests, as well as the percent of students with zero scores on each task, and mean scores for students with non-zero scores.

Figure 5. Summary of Performance on Kinyarwanda Literacy Subtests, P1
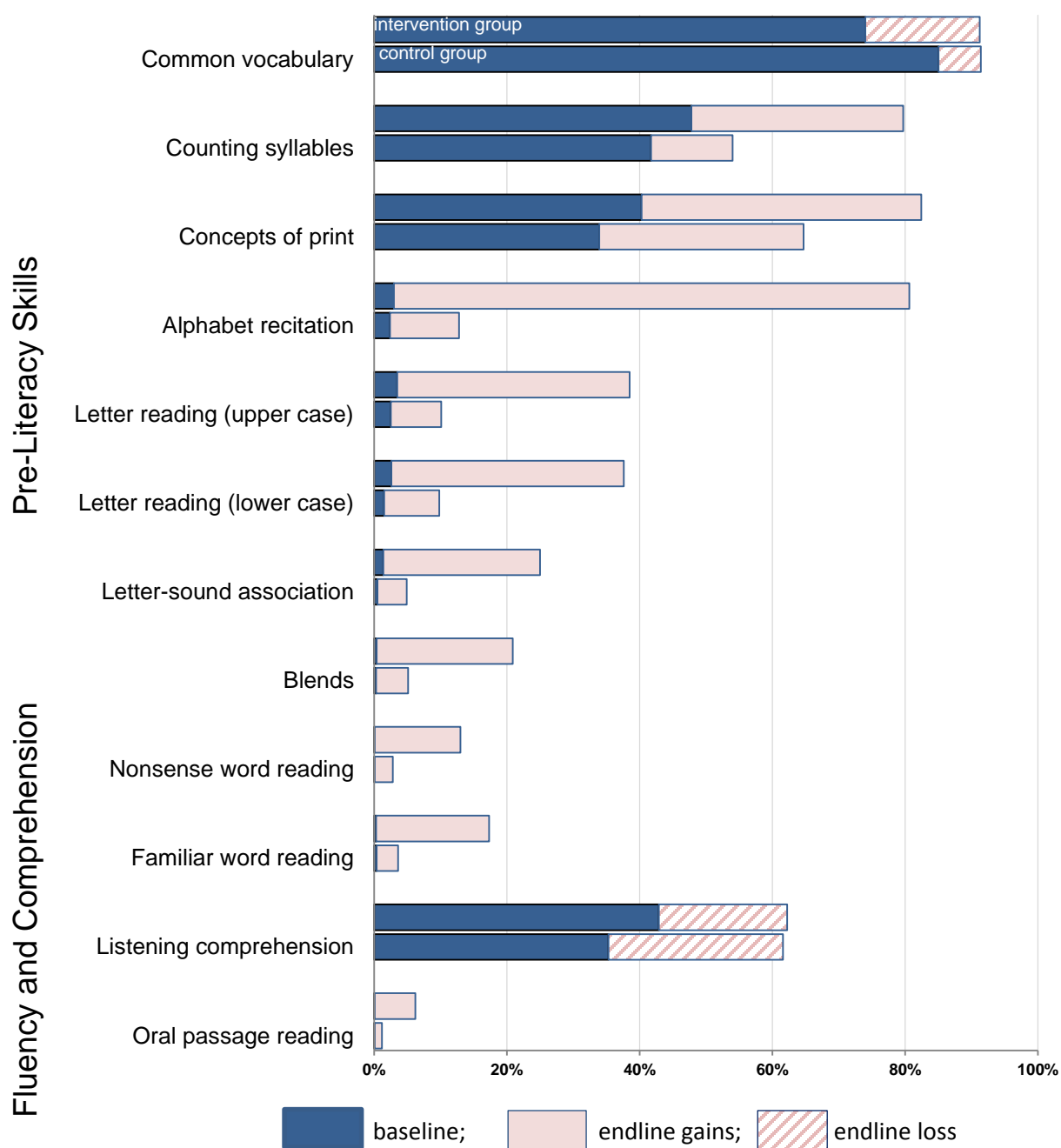
The comparison of means showed that the two study groups in P1 were equivalent at the baseline, and the intervention group did better than the comparison group on all but one subtests at the endline (at *p*<.001 level). Of the thirteen subtests in Kinyarwanda reading assessment, the statistical comparison of means showed that the two study groups in P2 showed that the two study groups were equivalent in three out of the twelve subtests, and in the remaining ten subtests the control group students scored higher at the baseline than the intervention group students. However, at the endline the intervention group scored statistically significantly higher than the control group in seven subtests.

Figure 6. Summary of Performance on Kinyarwanda Literacy Subtests, P2
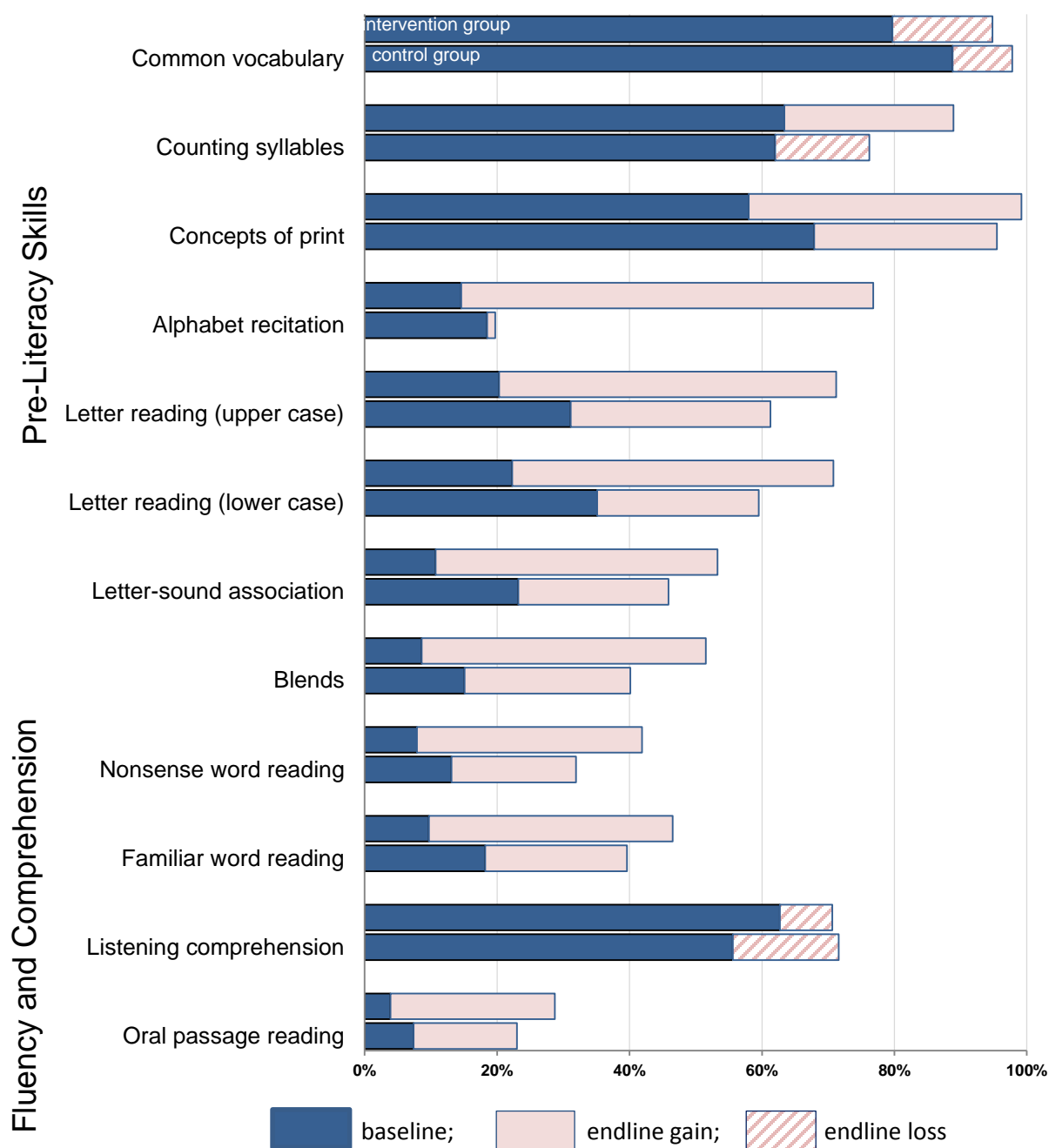
The table below shows a summary of the statistical tests of significance of difference in scores (independent group *t*-test) between the L3 students and the control group students at the endline.

Table 4. Summary results of the comparison of means between L3 and control group at the endline

		P1		P2	
		L3	**control**	**L3**	**control**
Pre-Literacy skills	Common vocabulary				)
	Alphabet recitation	sig. (*p*<.001)		sig. (*p*<.001)	
	Letter reading (upper case)	sig. (*p*<.001)		sig. (*p*<.05)	
	Letter reading (lower case)	sig. (*p*<.001)		sig. (*p*<.05)	
	Letter-sound association	sig. (*p*<.001)			
	Blends	sig. (*p*<.001)		sig. (*p*<.05)	
	Counting Syllables	sig. (*p*<.001)		sig. (*p*<.001)	
	Concepts of Print	sig. (*p*<.001)			
	Nonsense word reading	sig. (*p*<.001)		sig. (*p*<.05)	
Fluency and Comprehension skills	Familiar word reading	sig. (*p*<.001)			
	Listening comprehension	sig. (*p*<.001)		sig. (*p*<.05)	
	Oral passage reading	sig. (*p*<.001)			

Pre-Literacy Skills

Kinyarwanda pre-literacy skills of P1 and P2 students were tested using a range of subtests, including the following:

- Understanding common vocabulary words
- Knowledge of the alphabet
- Reading upper case and lower case letters
- Letter-sound association
- Reading blends
- Counting syllables
- Concepts of print

The following sections present assessment results by subtest.

Common Kinyarwanda Vocabulary Words. The first subtest of the Kinyarwanda literacy assessment gave students twenty instructions to perform simple actions that used basic vocabulary, such as "Point to your arm" and "Put the pen on the book". The results are somewhat surprising since both intervention and control group students did worse at the endline compared to the baseline. The graph below shows that an average percent of correct answers went down from over 90 percent to 75 for intervention group and 85 for control group P1 students, and from almost 100 percent to 80 and 90 percent for intervention and control P2 students.

Data analysis did not reveal any substantial differences in performance between boys and girls on this subtest.

Kinyarwanda Alphabet Knowledge. Figure 7 shows the results of the Kinyarwandan alphabet recitation subtest. At the baseline, fewer than one in twenty P1 students in either group were able to recite some letters of the alphabet, and only about one in five P2 students were able to do so. At the endline, intervention students showed a dramatic progress, with an average of 80 percent of alphabet recited correctly by students from both grades.

Figure 7. Mean Percent of Kinyarwanda Alphabet Recited Correctly, by Grade
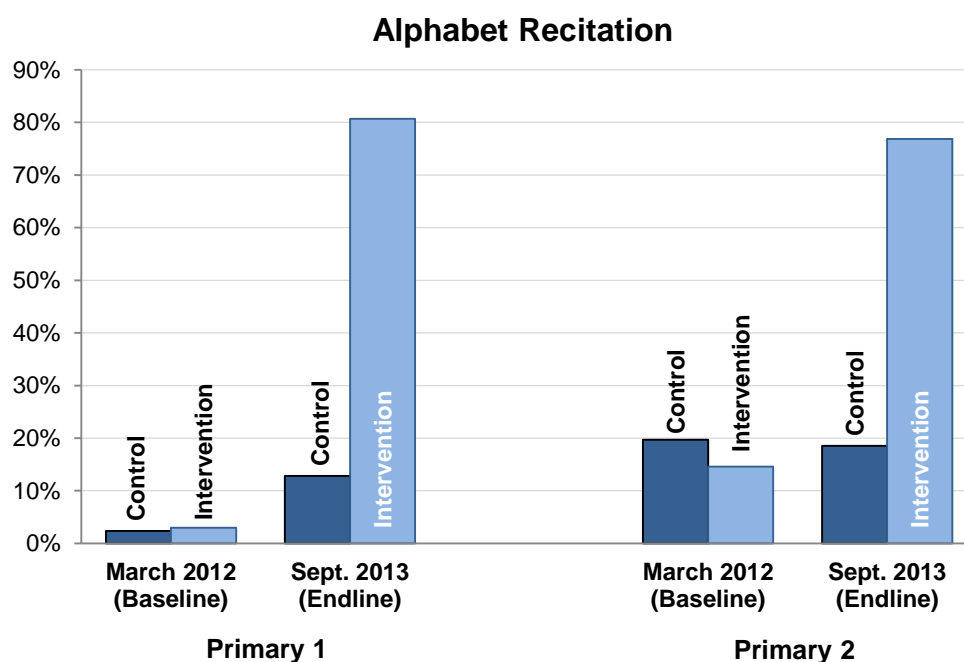
Figure 8 shows the distribution of students in groups corresponding to their achievement on this test. The vertical axis shows the percent of students, and the horizontal axis shows the percent correct scored on this subtest. These figures show how most P2 intervention students moved from scoring zero or below 20 percent correct at the baseline to reciting the alphabet correctly at the endline. Control group students did not show such gains. There were no substantial differences in performance between genders.

Figure 8. Distribution of Percent of Alphabet Recited Correctly by P2 Students

P2 Baseline

P2 Endline
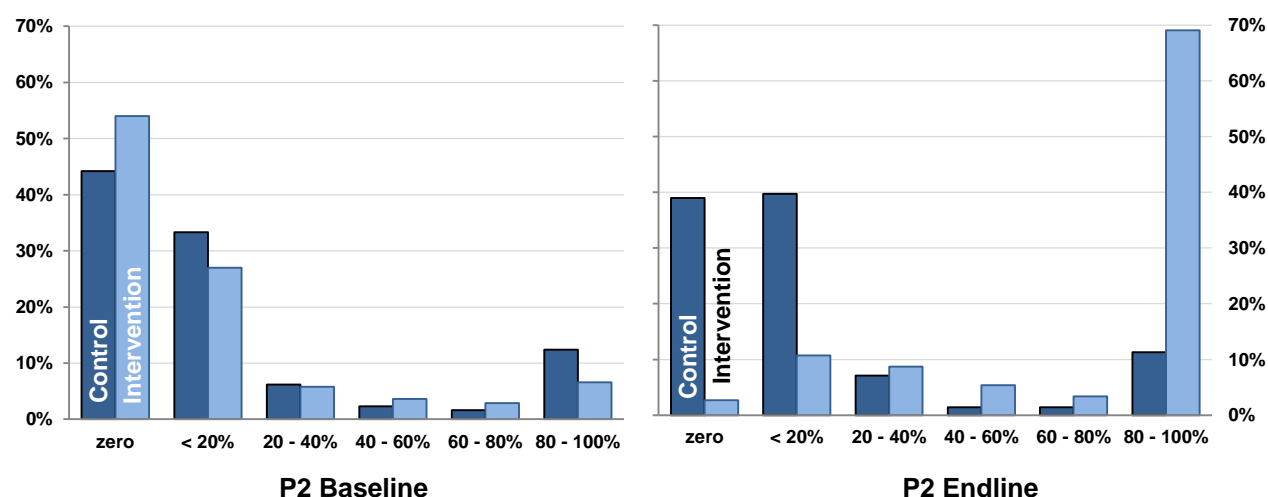
Upper and Lower Case Letter Reading. Students were also asked to read the letters of the alphabet, presented out of order. First, they were asked to read upper case letters, and then lower case letters. As Figure 10 demonstrates, intervention group students in both grades improved more between the two measurements in this subtest. This improvement is particularly evident among P2 students, where intervention group students started out lower than control group students, but did better than control group counterparts at the endline.

Figure 9. Mean Percent of Upper Case and Lower Case Alphabet Read Correctly, by Grade

Upper Case Alphabet Reading

Lower Case Alphabet Reading
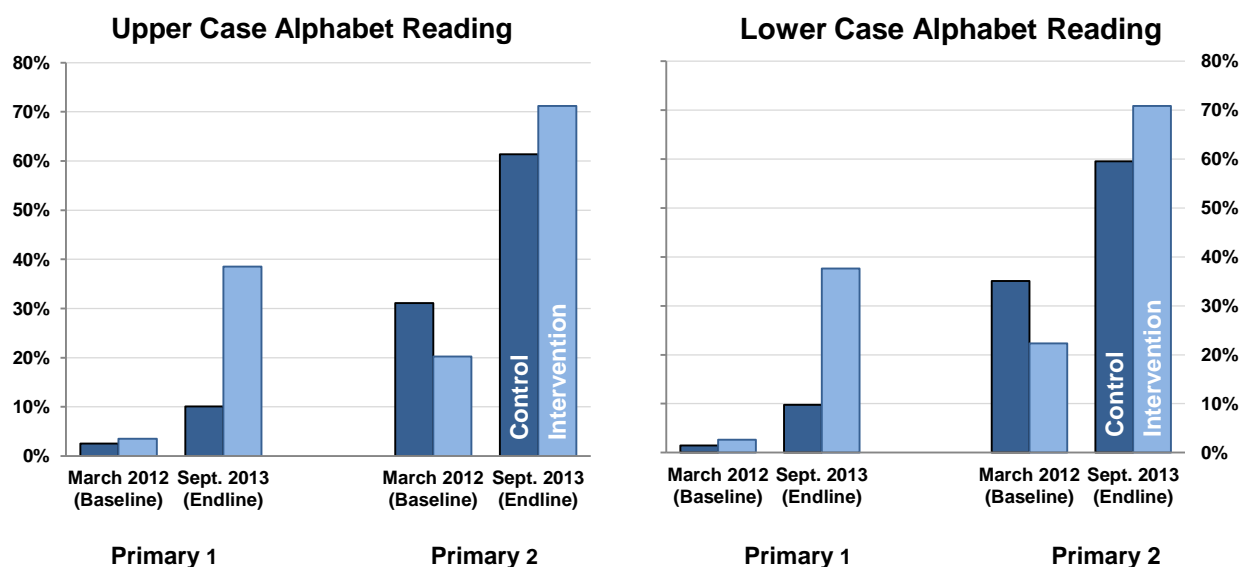
The proportion of control group students with zero scores on the upper case alphabet reading subtest dropped a little between the baseline and the endline, from 28 percent to 19 percent. However, the proportion of intervention group students with zero scores on this subtest dropped from 35 percent to 9

percent. A similar trend is observed with the lower case alphabet reading subtest. There was no substantial difference between boys and girls in their performance on these two subtests.

Letter-Sound Association. The next subtest in this section presented students with 100 letters, a mix of lower case and upper case, which they were supposed to read within 90 seconds. The test results show that while both groups improved between the baseline and the endline, the intervention group students gained significantly more. In traditional Kinyarwanda literacy instruction, teachers typically focus on blends, not individual letters, so it is of little surprise that the scores on this subtest were low, particularly among control group students.

Figure 10. Mean Percent of Correct Letter-Sound Association, by Grade
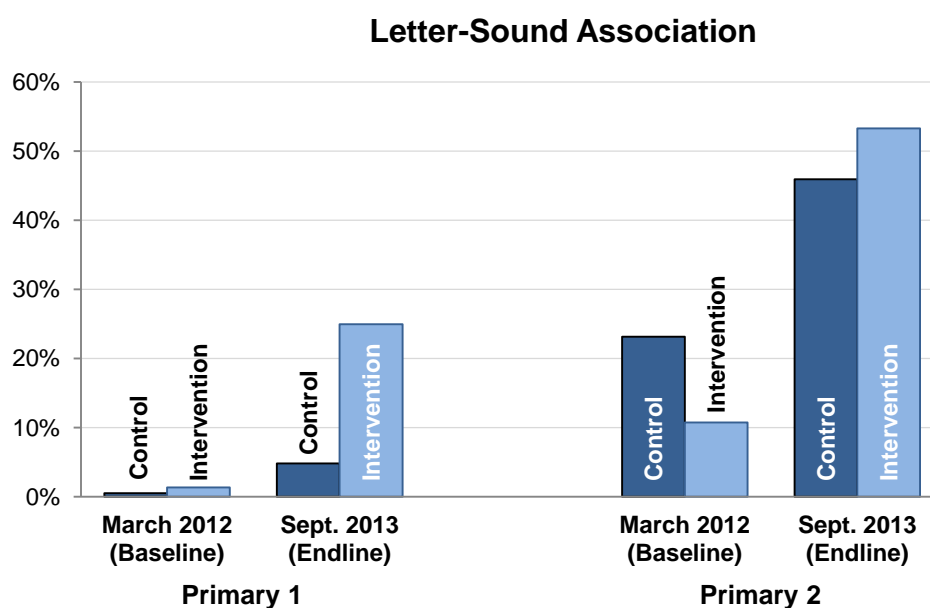
As the distribution graphs below show, a large proportion of learners scored zero on this task. Among intervention group students, a proportion with zero score dropped by half between the baseline and the endline, but among control students it only dropped by a few percentage points. The overall distribution is U-shaped, with more students scoring zero and 80 to 100 percent than in the midrange.

Figure 11. Distribution of Percent of Correct Letter-Sound Association by P2 Students

P2 Baseline

P2 Endline
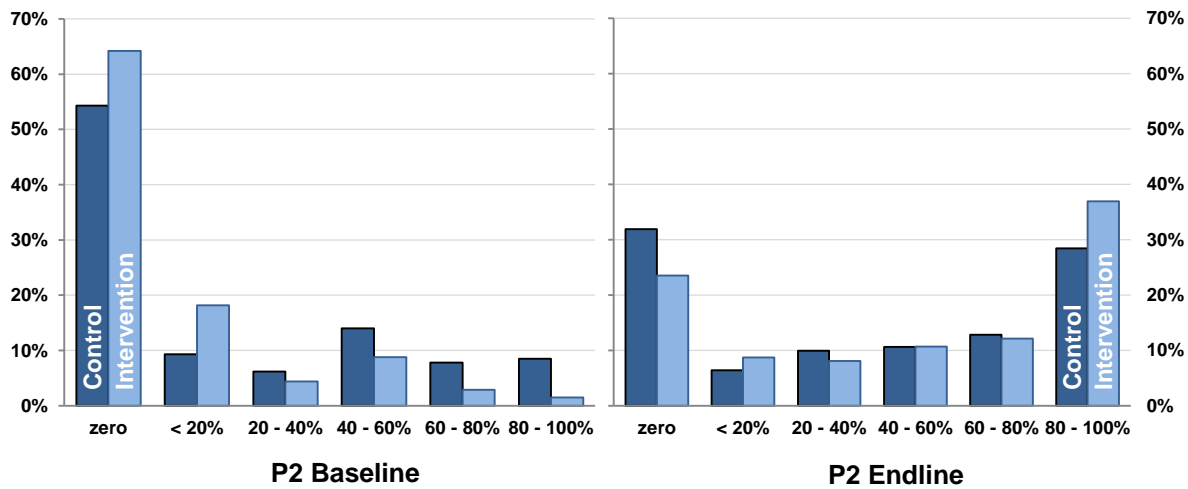
Gender comparisons did not find any substantial differences between boys and girls on this subtest.

**Reading Blends**. Certain combinations of two or more letters are called blends. Letter blends can appear in the beginning, middle, or at the end of words to create specific sounds. Ability to read blends is an important foundational skill that helps early readers gain fluency. In L3 Kinyarwanda literacy assessment, students were asked to read 100 blends within the allocated time of 90 seconds. At the baseline, almost no P1 students could read blends in the test, and fewer than one in five P2 students could read them. At the endline, the average percent correct was 20 for P1 intervention students, and just over 50 percent for the P2 intervention students. P2 control group scored on average 10 percent lower at the endline.

The distribution of percent of blends read correctly is U-shaped, similar to the letter-sound association one. This distribution shows that most learners

Teacher-made blend poster

20
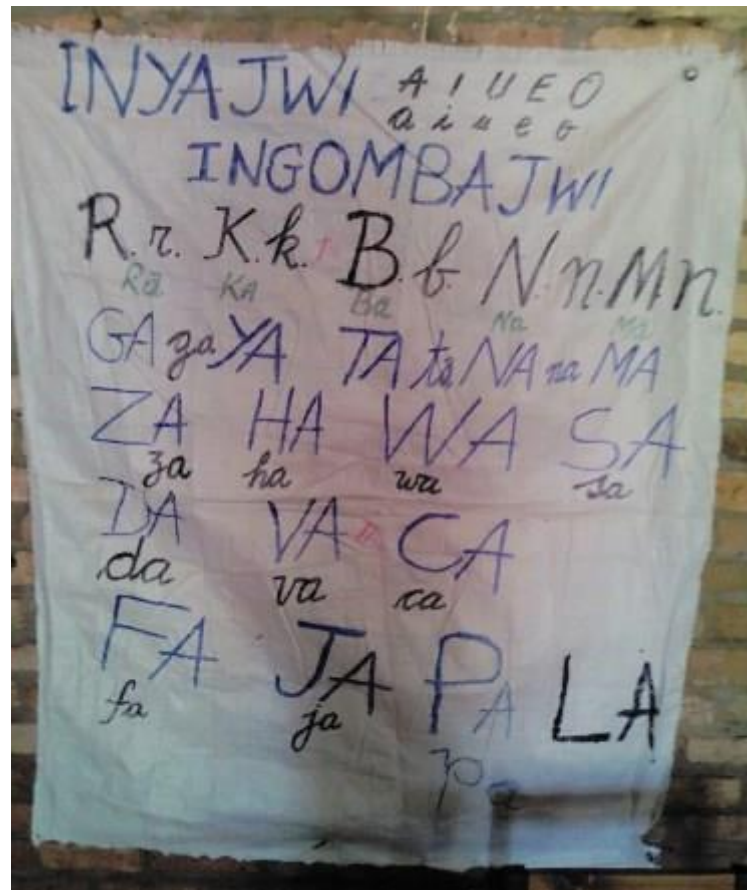
fall at either end of the spectrum, either not reading any blends correctly, or reading all or nearly all of them correctly. Nearly a quarter of intervention P2 students, and almost 40 percent of control P2 students did not read a single blend correctly at the endline.

Figure 12. Distribution of Percent of Blends Read Correctly by P2 Students

Gender comparisons did not show any substantial differences between boys and girls on this subtest.

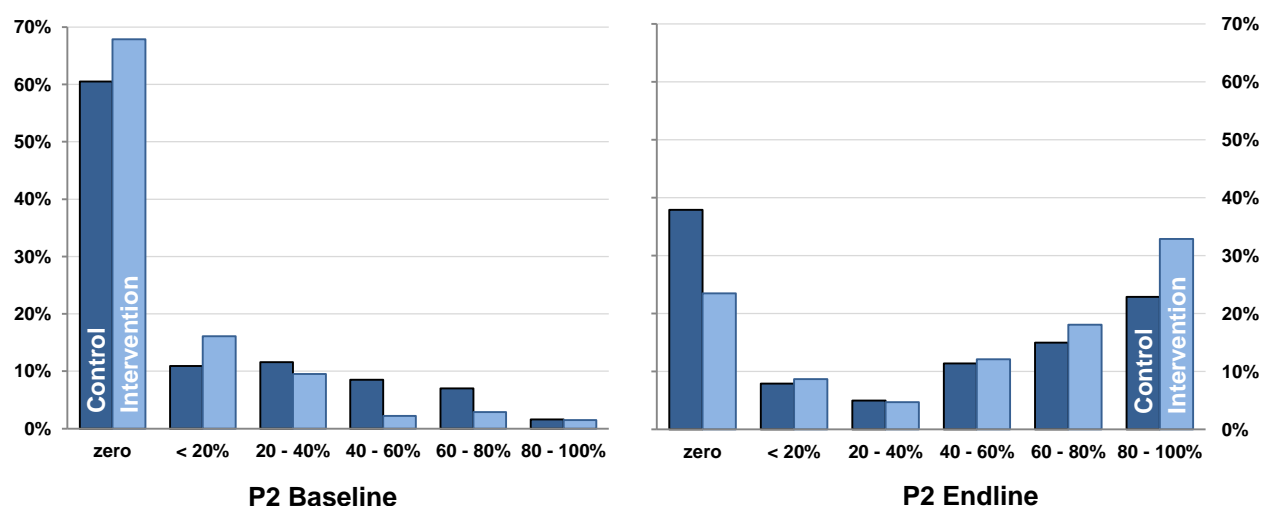**Counting Syllables.** The counting syllables subtests included 10 items, ranging from monosyllabic to words with three syllables. Students were asked to identify how many syllables each word had. As the distribution below shows, mean scores for this subtest were higher than for other subtests in the assessment. On average, P1 students were able to count correctly about half of the syllables at the baseline, and P2 students were able to count about 60 percent correctly. At the endline, intervention group students demonstrated significant gains: on average, P1 intervention students counted syllables correctly in 8 out of 10 words, and P2 intervention students counted syllables correctly in 9 out of 10 words.

Figure 13. Mean Percent of Syllables Counted Correctly, by Grade
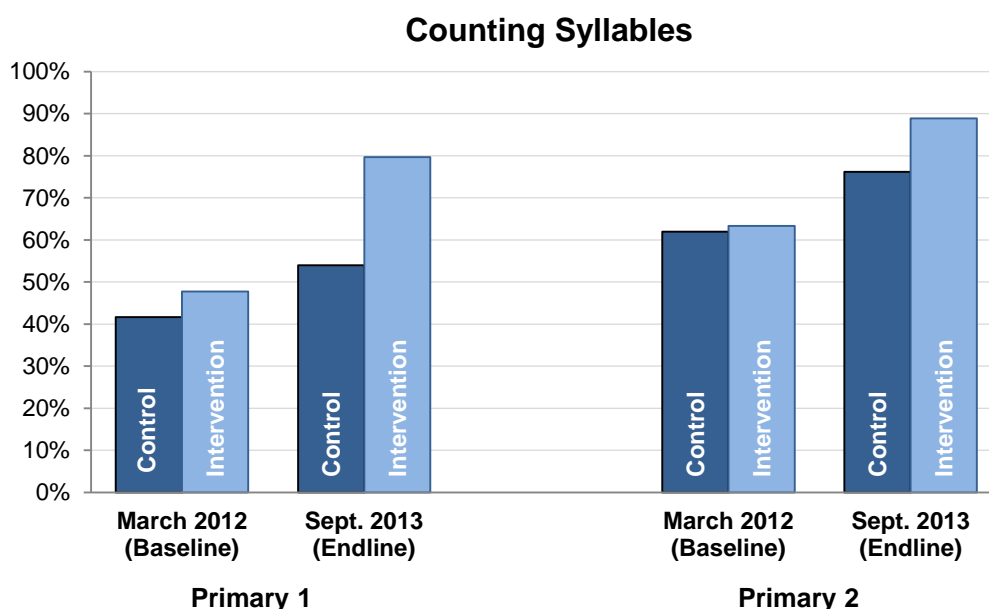
About a quarter of P2 students had zero scores at the baseline, but the majority of P2 students were able to count syllables correctly in 80 to 100 percent of words. In intervention group, the number of students with zero scores on this task dropped almost to zero at the endline. There were no discernable differences in the performance on this subtest between boys and girls.

Concepts of Print. Concept of print subtest measures students' familiarity with the conventions of text, and asks the following questions:

- "Point to the cover of the book"
- "Where would you begin to read? Show me with your finger"
- "Show me in which direction you would read the text"
- "Show me the word 'zebra'"
- "Show me the word 'crocodile'"

As the graph below shows, the majority of students did well on this subtest. P1 students started at about 35 to 40 percent correct at the baseline and reached 56 (control) to 70 (intervention) percent correct at the endline. P2 students also improved, particularly the intervention group students whose scores rose from just under 60 percent correct at the baseline to over 80 percent correct at the endline.

Figure 14. Mean Percent of Concepts of Print Performed Correctly, by Grade
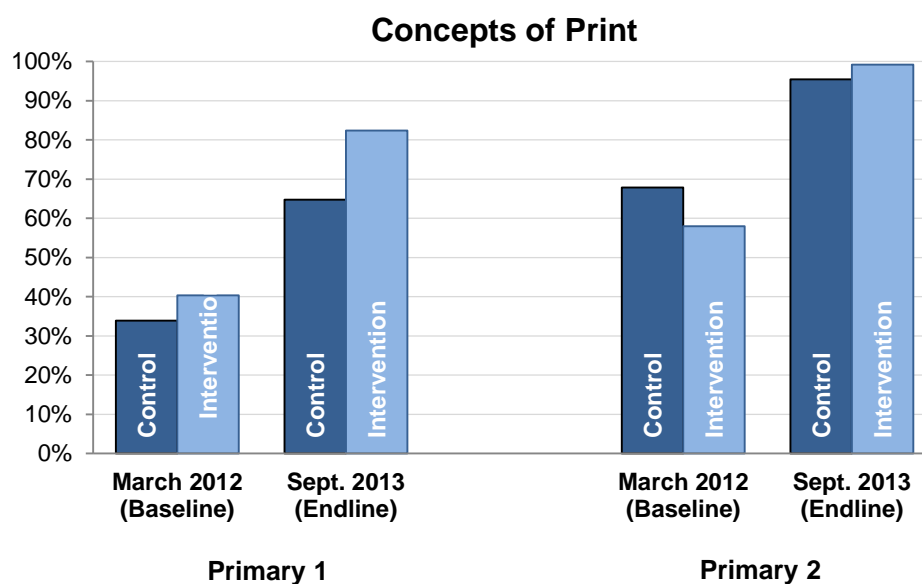
Boys did better on this subtest than girls. In both intervention and control group a much higher proportion of boys than girls answered all or almost all questions correctly. The graph below shows the distribution of the results among P2 students at the endline.

Figure 15. Distribution of Percent of Concepts of Print Performed Correctly by P2 Students at the Endline, by Gender
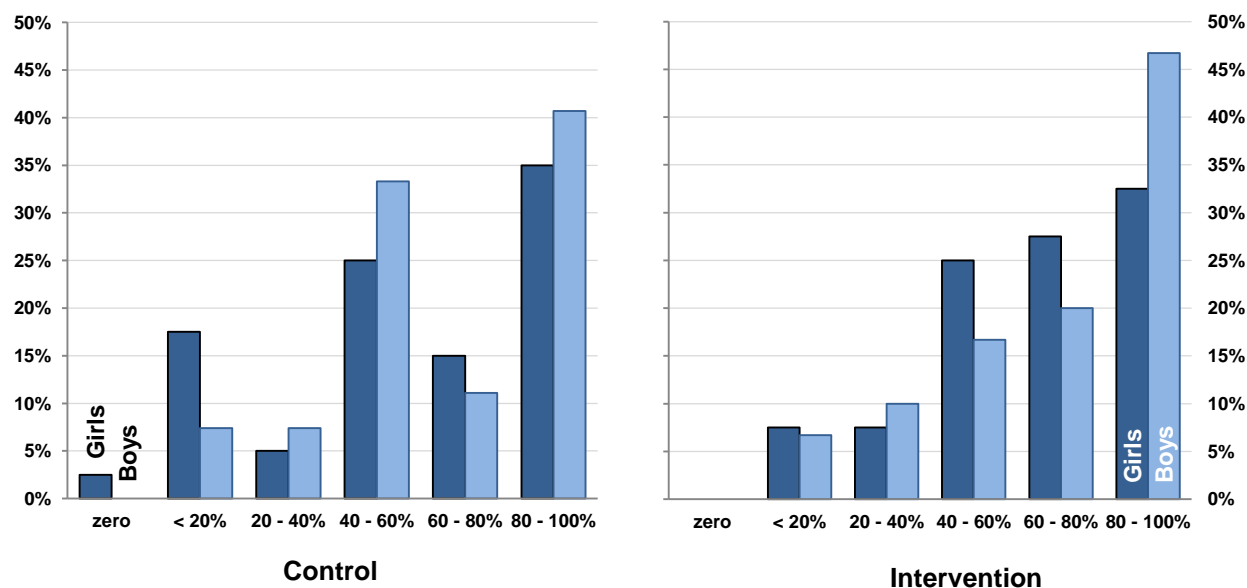
Fluency and Comprehension Skills

Kinyarwanda literacy assessment included four subtests designed to test students' ability to read and understand text:

- Reading nonsense words
- Reading familiar words
- Reading a connected text passage
- Listening comprehension

Nonsense Word Reading. On the simple nonsense word (non-word) decoding, students were asked to read 50 invented words, within 90 seconds. This subtest measures students' ability to apply the rules of decoding to unfamiliar words, and is different from reading familiar words, which are frequently memorized and then recognized in the text, instead of being decoded.
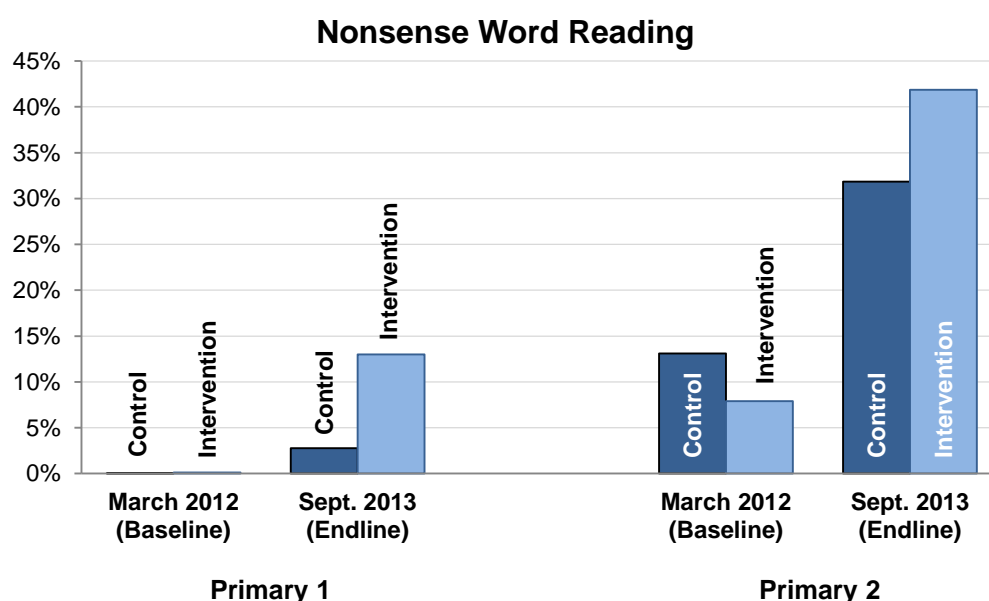
The results of the analysis show that intervention group students made a much larger progress between the baseline and the endline when compared to the control group students. Although the difference is dramatic, students in P1 in particular show an overall low achievement in this subtest.

Figure 16. Mean Percent of Nonsense Words Read Correctly, by Grade

The results of the assessment show that the majority of students could not decode a single word. At the baseline, nearly 70 percent of control group P2 students, and over 75 percent of intervention P2 students had zero scores. At the endline, the proportion of intervention students with zero scores was reduced by more than half – from 75 percent to less than 30, and the proportion of control students with zero scores dropped form 63 percent to 45 percent. These results demonstrate that although more students were able to decode nonsense words at the endline, there is still a very sizable proportion of students who entirely fail to grasp principles of decoding.

Familiar Word Reading. Familiar word reading subtest measures speed and accuracy of reading commonly used words. Students may know some or all of the test words by sight, or be able to decode them. Differing from the nonsense word section, all words in the familiar word section are meaningful and frequently used.

At the baseline, P1 students were not able to read the words from the list, and P2 students were able to read less than 20 percent of words accurately and within the allocated time of 90 seconds. Both intervention and control group students made substantial gains between the baseline and the endline, but intervention students gained significantly more, as the graph below demonstrates.

Figure 17. Mean Percent of Familiar Kinyarwanda Words Read Correctly, by Grade
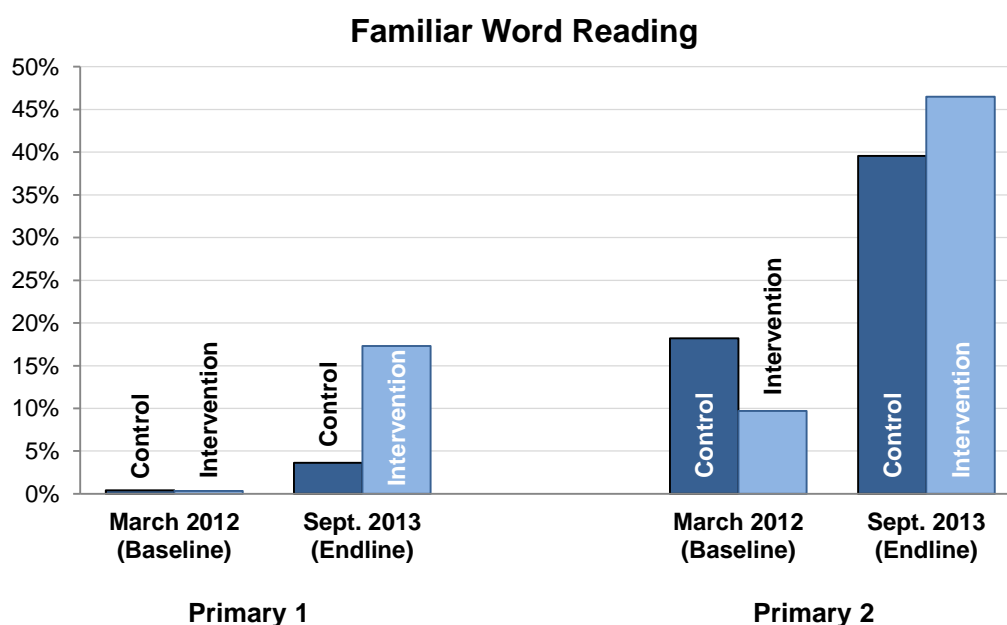
The P2 baseline/endline comparison graph below shows that the proportion of students with zero scores dropped from about 50 to 35 percent in the control group, and from almost 70 percent to 24 percent in the intervention group. While intervention group students demonstrated overall lower scores at the baseline, their scores at the endline were higher than their counterparts from the control group.

Overall results for this section are rather low, with just about one in four to one in five students reading 80 to 100 percent of familiar words from the list correctly.

Figure 18. Distribution of Percent of Words Read Correctly by P2 Students

P2 Baseline

P2 Endline
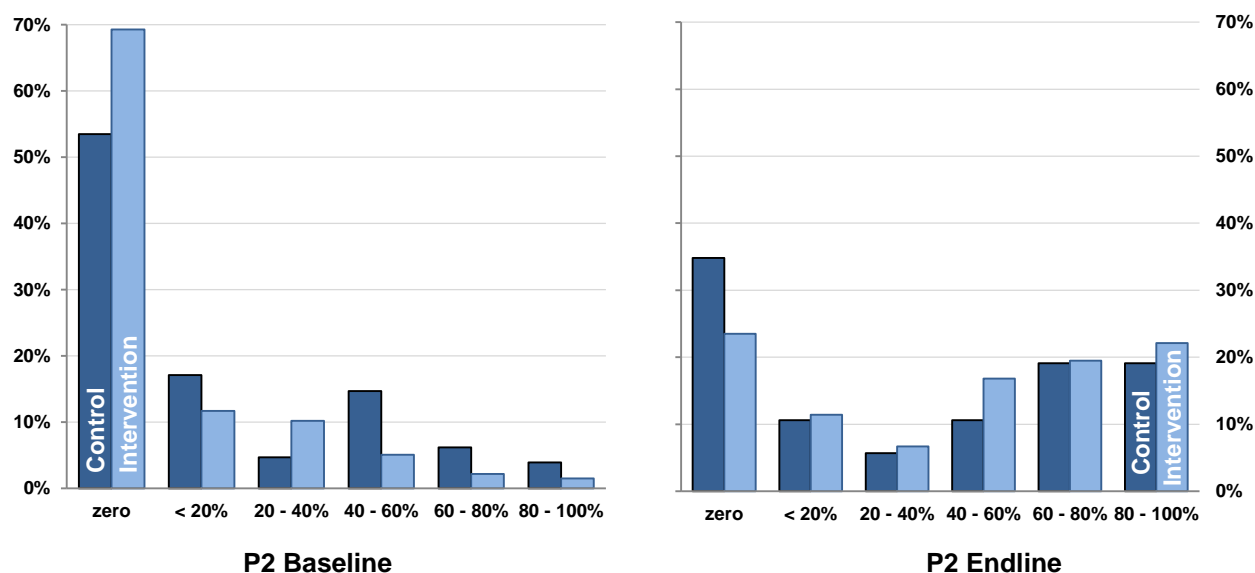
Listening Comprehension and Kinyarwanda Vocabulary. On the **listening comprehension subtest**, assessors read a passage to student and asked him/her eleven comprehension questions.   The graph below shows the distribution of the mean correct answers for both control and intervention groups. Both groups had a decrease in mean scores between the baseline and the endline, and the control group showed a larger decrease, compared to the intervention group.

Figure 19. Mean Percent of Correct Answers on Listening Comprehension in Kinyarwanda Subtest, by Grade

Listening Comprehension
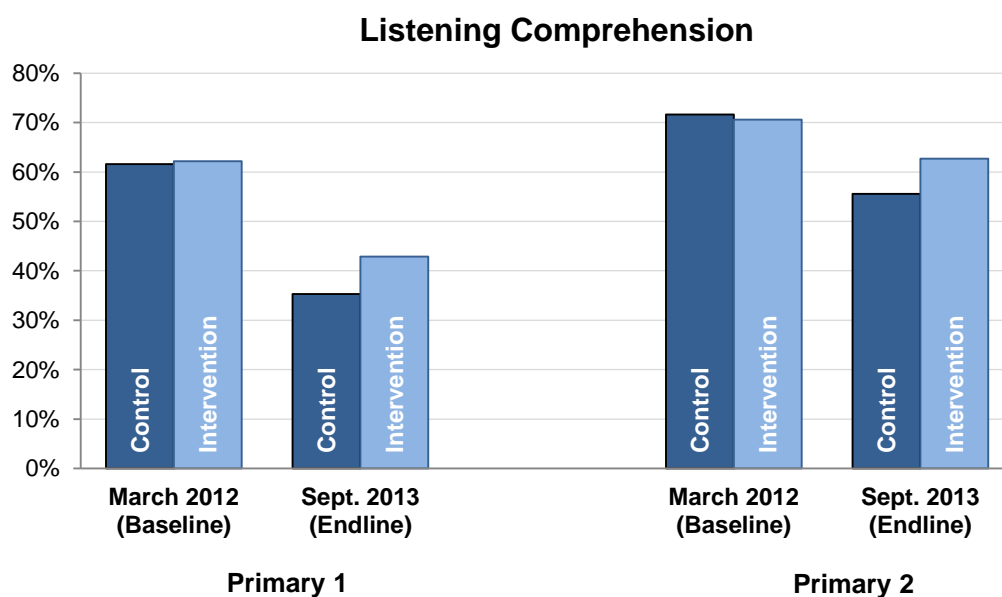
Oral Passage Reading in Kinyarwanda. On the passage reading and comprehension, students were scored on the words they read correctly in the passage (total possible 76). Students had a time limit of 120 seconds to read the text. The results presented in the graph below show that very few P1 students were able to read even a single word in the text. Although the endline P1 results are better than the baseline, the mean percent of words read correctly is still very low: just over 1 percent for the control group, and about 6 percent for the intervention group.
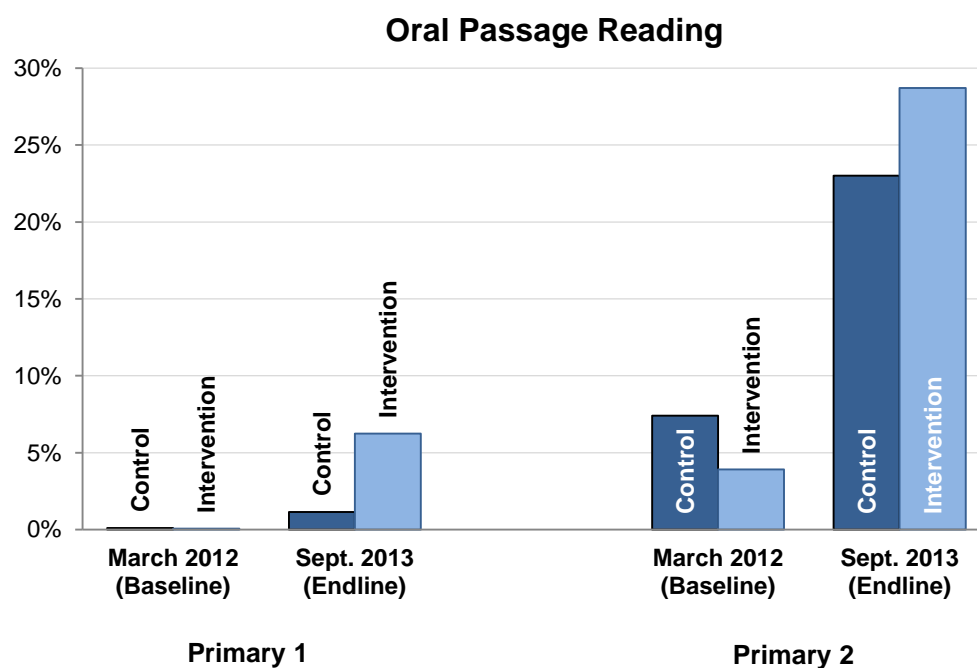
P2 students did better on this test, and also improved significantly more between the baseline and the endline. However, the mean results at the endline were still low, with just about 25 percent of words read correctly.

Figure 20. Mean Percent of Words in a Passage Read Correctly, by Grade

The baseline/endline comparison graph below shows that the proportion of students with zero scores went down substantially for both groups, but particularly for the intervention group students: from over 70 percent at the baseline to just over 30 percent at the endline. While the baseline distribution was heavily skewed toward zero scores, the endline scores are more normally distributed, showing a growth in both the middle scores area and even in the 80 to 100 percent quintile.

Figure 21. Distribution of Words in a Passage Read Correctly by P2 Students

P2 Baseline

P2 Endline
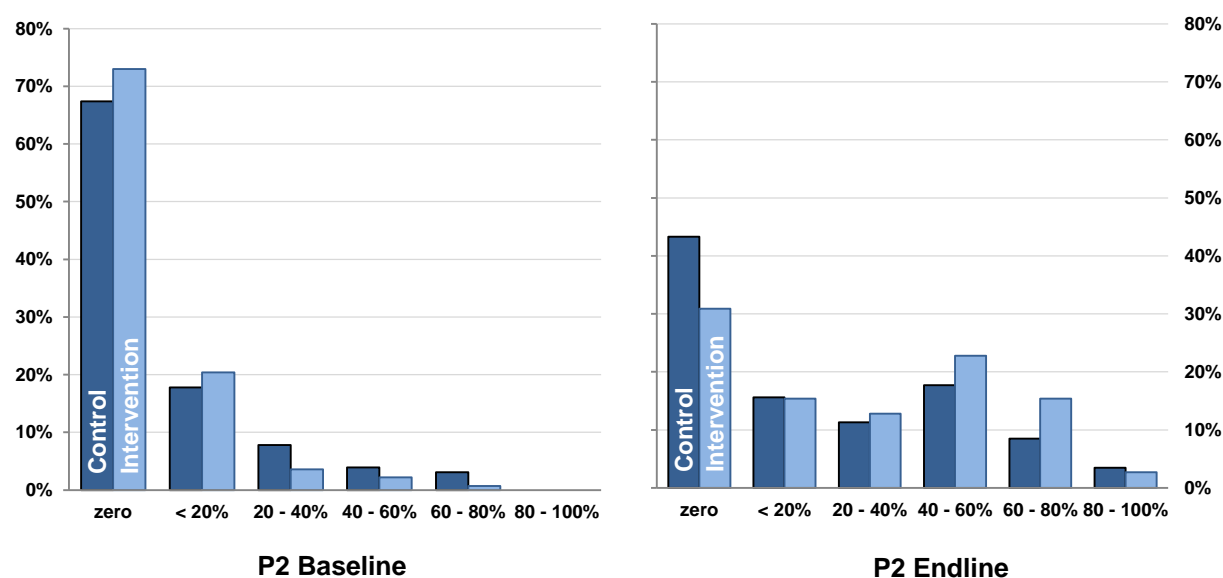
An important predictor of successful reading is fluency – the measure of speed of accurate reading. Fluency is expressed in words correct per minute and is computed by dividing the total number of words read correctly by the seconds it took to read them, and then multiplying by 60 seconds.

Fluency of tested students improved between the baseline and the endline, although overall remained relatively low. P1 intervention students read correctly only 5 words a minute at the endline. P2 intervention students achieved fluency of about 21 words per minute at the endline, up from barely 2 at the baseline. Control group P2 students' fluency went up from 3 to 17 words per minute, as the graph below demonstrates.

Figure 22. Mean Words Correct Per Minute, by Grade
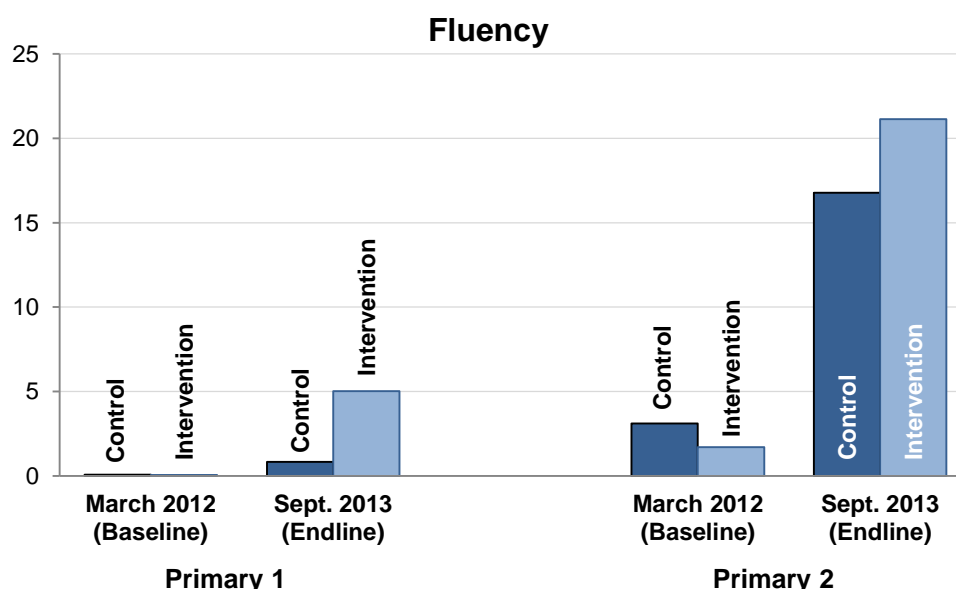
Since fluency benchmarks for P2 have not been established yet, the study does not draw conclusions about a proportion of students meeting proficiency standards.

Summary of Findings

P1 students in both intervention and control study groups demonstrated moderate skill level in three out of twelve Kinyarwanda literacy assessment subtests:

- Basic vocabulary
- Alphabet recitation
- Counting syllables

P1 students scored over 50 percent, on average, on these three subtests at the endline. Proficiency in these three subtests ensures mastery of foundational pre-literacy skills that are necessary for reading later on. Overall, intervention students showed statistically significantly larger gains between the pretest and the posttest in all but one subtest: basic vocabulary.

P2 students scored on average over 50 percent correct on the following subtests:

- Basic vocabulary
- Alphabet recitation
- Alphabet reading (both upper case and lower case)
- Letter-sound association (intervention group only)
- Reading blends (intervention group only)
- Counting syllables
- Concepts of print
- Listening comprehension

These subtests measure pre-literacy skills. P2 students, on average, scored below 50 percent on all subtests aimed at measuring actual reading fluency, such as decoding unfamiliar words, reading familiar words, and reading a connected text.

The examination of the assessment results revealed a high proportion of students with zero scores in individual subtests. Further data analyses showed that scores in many of subtests are distributed in a U-shape, with a high proportion of zero scores and with high scores, and with only few students achieving medium results. These results portray a diverse classroom, filled with students with different ability levels and in need of different instructional approaches. While students with zero scores might require remediation, students with high scores need leveled readers and enhanced opportunities to practice their reading skills.

Data analysis showed that the intervention was particularly effective for P1 students who scored significantly higher on the endline assessment, compared to the control group students. P2 intervention students also scored higher than their control counterparts, but the difference between the two study groups was less pronounced.

ENGLISH ASSESSMENT FINDINGS

Sampled P1 and P2 students were tested in English literacy skills, using an adapted Early Grade Reading Assessment. As Figure 32 illustrates, the overall level of English reading proficiency is low. Students from intervention schools from both P1 and P2 grades demonstrated some gains between the baseline and endline assessments, while the scores of the control group students remained flat.

Figure 23. Overall Scores[3] English, by Grade

Overall Scores in English
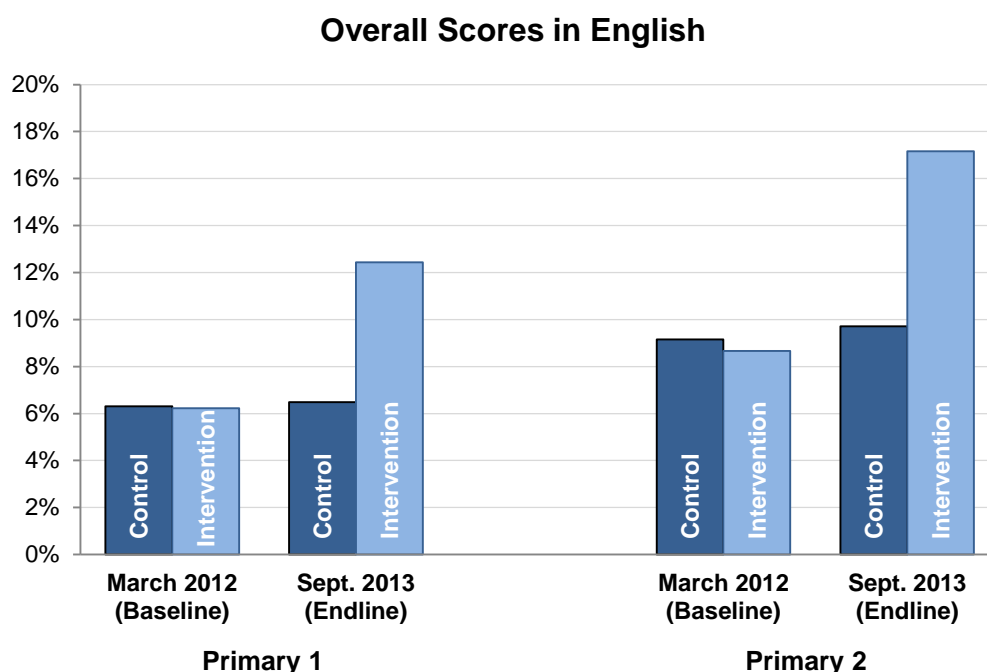
Annex 3 shows the mean distributions for all English literacy subtests, as well as the percent of students with zero scores on each task, and mean scores for students with non-zero scores at the baseline and the endline. Annex 6 contains additional analyses.

The figure below shows the distribution of the average scores at the baseline and the endline for P1 students, for each of the English literacy assessment subtests. The figure shows the average percent correct achieved by P1 students at the baseline (blue shading) and the gain made at the endline (pink shading). The subtests where gains are shown in stripes are those where students scored lower at the endline than at the baseline.

The comparison of means showed that the two study groups in P1 were equivalent at the baseline on all but one subtest (alphabet recitation) on which the control group students did better. The intervention group did better than the comparison group on all subtests at the endline.

---

[3] The values in the graph are means of all subtest means.

Figure 24.Summary of Performance on English Literacy Subtests, P1
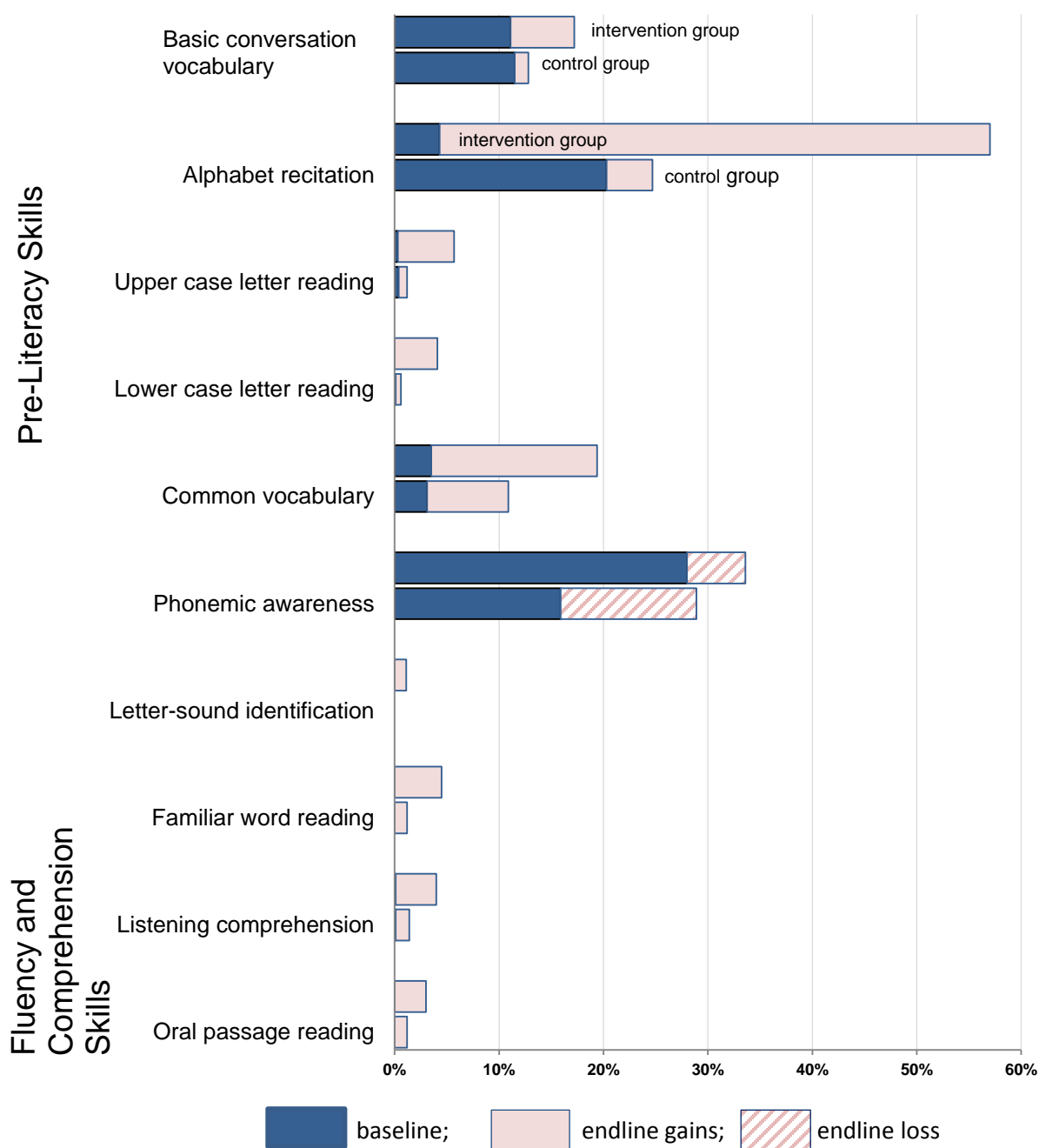
Of the ten subtests in English reading assessment, the statistical comparison of means showed that the two study groups in P2 were equivalent in all but two subtests (basic conversation vocabulary and alphabet recitation. However, at the endline the intervention group scored statistically significantly higher than the control group in seven out of ten subtests.

Figure 25.Summary of Performance on English Literacy Subtests, P2
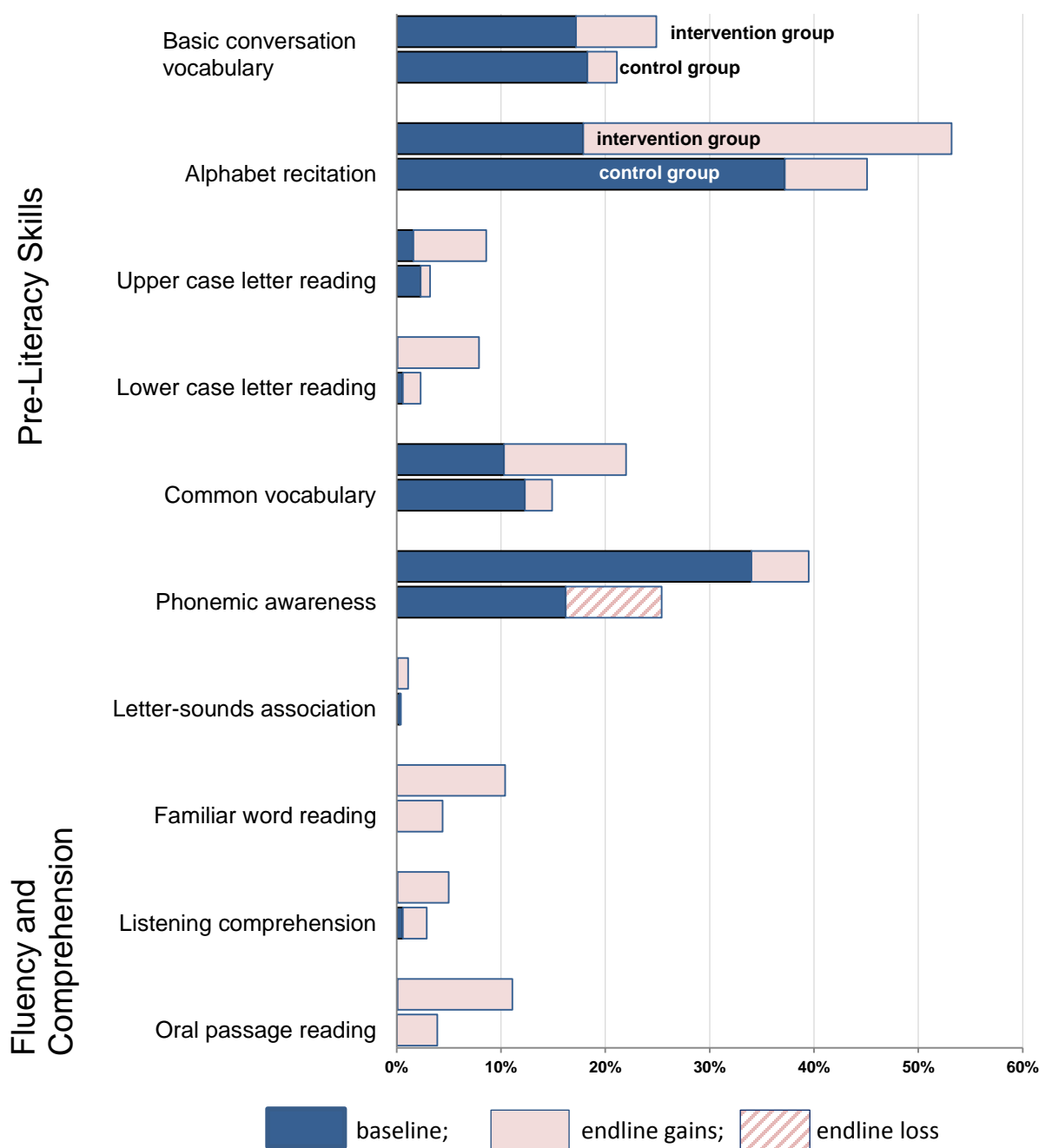
The table below shows a summary of the statistical tests of significance of difference in scores (independent group *t*-test) between the L3 students and the control group students at the endline. As the table shows, the intervention group students achieved statistically significantly larger scores than the control group students in almost all subtests.

Table 5. Summary results of the comparison of means between L3 and control group at the endline

|		P1		P2	
		L3	**control**	**L3**	**control**
Pre-Literacy skills	Basic Conversation	sig. ($p<.001$)		sig. ($p<.001$)	
	Alphabet recitation	sig. ($p<.001$)			
	Letter reading (upper case)	sig. ($p<.001$)		sig. ($p<.001$)	
	Letter reading (lower case)	sig. ($p<.001$)		sig. ($p<.001$)	
	Phonemic awareness	sig. ($p<.001$)		sig. ($p<.001$)	
	Letter-sound association	sig. ($p<.001$)			
	Common Vocabulary	sig. ($p<.001$)		sig. ($p<.05$)	
Fluency and Comprehension skills	Familiar word reading	sig. ($p<.001$)		sig. ($p<.001$)	
	Listening comprehension	sig. ($p<.05$)			
	Oral passage reading	sig. ($p<.05$)		sig. ($p<.001$)	

Pre-Literacy Skills

English pre-literacy skills of P1 and P2 students were tested using a range of subtests, including the following:

- Understanding basic vocabulary words used in a conversation
- Knowledge of the alphabet
- Common vocabulary
- Reading upper case and lower case letters
- Letter-sound association
- Phonemic awareness

The following sections present assessment results by subtest.

Basic Conversation Vocabulary. The first subtest of the assessment involved the assessor asking the student ten questions involving basic vocabulary, such as: "What is your name?", "What do you like to do?", and "What day is it today?" At the baseline, P1 students answered on average only one question out of ten correctly, and P2 students answered two out of ten questions correctly. At the endline, intervention students showed a 30 to 50 percent improvement, although the overall scores remained low.

Figure 26. Mean Percent of Conversational Skills Tasks, by Grade

Basic Conversation Vocabulary
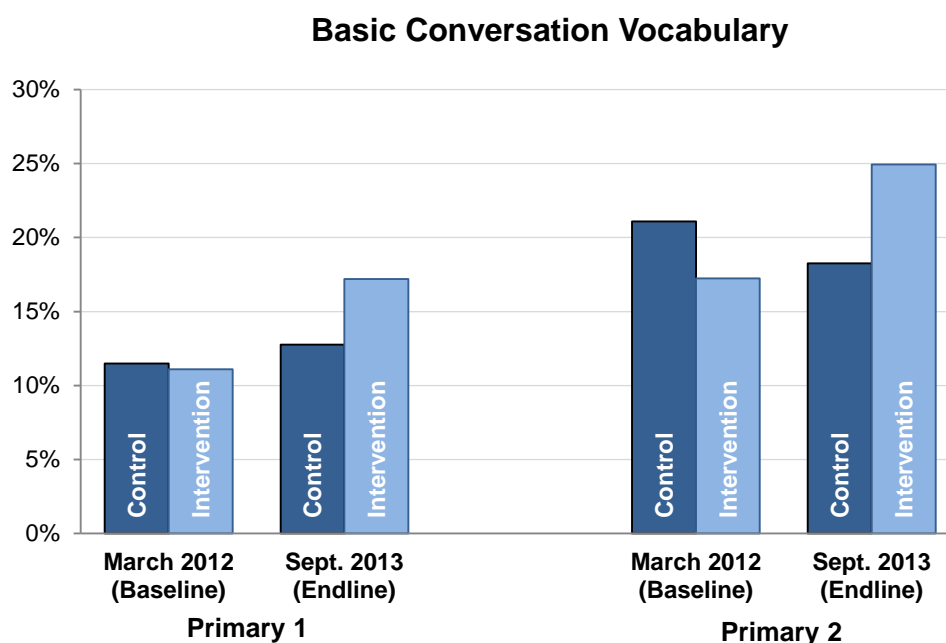
The grouped distributions show that a vast majority of students were able to answer only one or two question, and very few students answered more than two questions. About one in five students did not answer a single question correctly. Data analysis did not find any substantial differences in the performance on this subtest between boys and girls.

Common Vocabulary Words. A subtest of the English literacy assessment asked students twenty questions that used basic vocabulary, such as "Show me your head" and "Put the pen on the book". The results in the summary graph below show that the overall level of students' familiarity with basic vocabulary words is quite low, although both P1and P2 students showed substantial improvement between the baseline and the endline. Still, at the endline, control P2 group students answered only 3 questions out of 20 correctly, and intervention P2 students answered between 4 and 5 questions out of 20 correctly.

Teacher-Made English Vocabulary Materials

Figure 27. Mean Percent of Common Vocabulary Words Read Correctly, by Grade
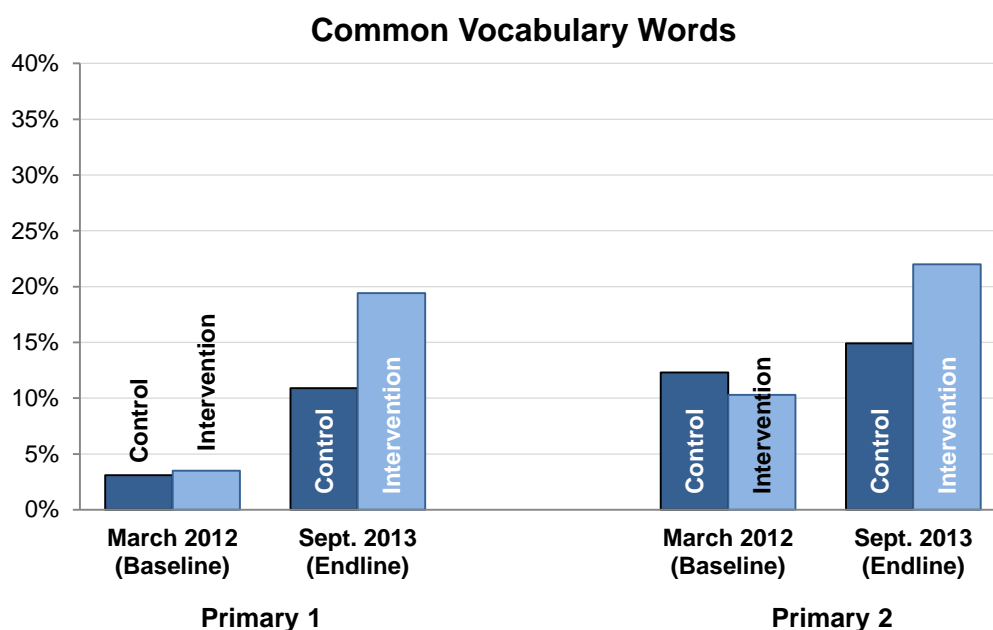
Data analysis showed that the proportion of students with zero scores went down substantially between the baseline and the endline, particularly among the intervention students. No substantial differences between genders were observed.

Figure 28. Distribution of Percent of Common Vocabulary Words Read Correctly by P2 Students
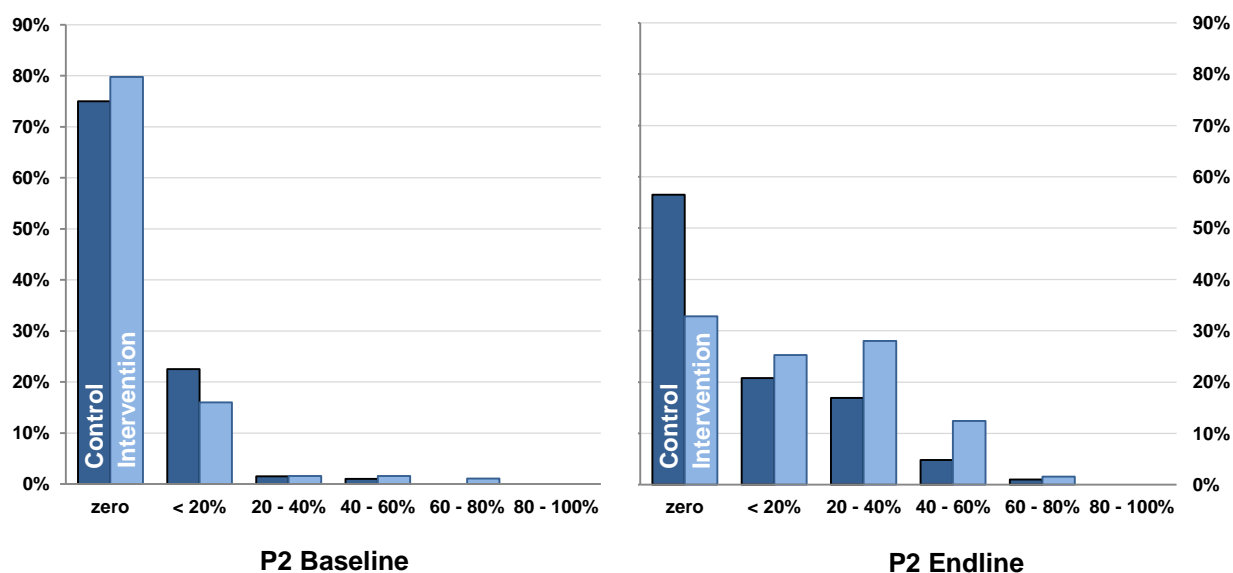
Alphabet Knowledge The second section of the English assessment aimed to assess students' knowledge of English alphabet. It included the following subtests:

- English alphabet recitation

- Reading of upper case letters of the alphabet (letters presented out of order)
- Reading of lower case letters of the alphabet (letters presented out of order)
- Reading of a mix of upper case and lower case letters of the alphabet

All three subtests were timed at 60 seconds each.

The graph below shows that at the baseline, P1 and P2 students were not able to recite the English alphabet correctly. At the endline, intervention students showed a significant progress, with an average of 55 percent of the alphabet recited correctly by students from both grades.

Figure 29. Mean Percent of Alphabet Recited Correctly, by Grade

Alphabet Recitation
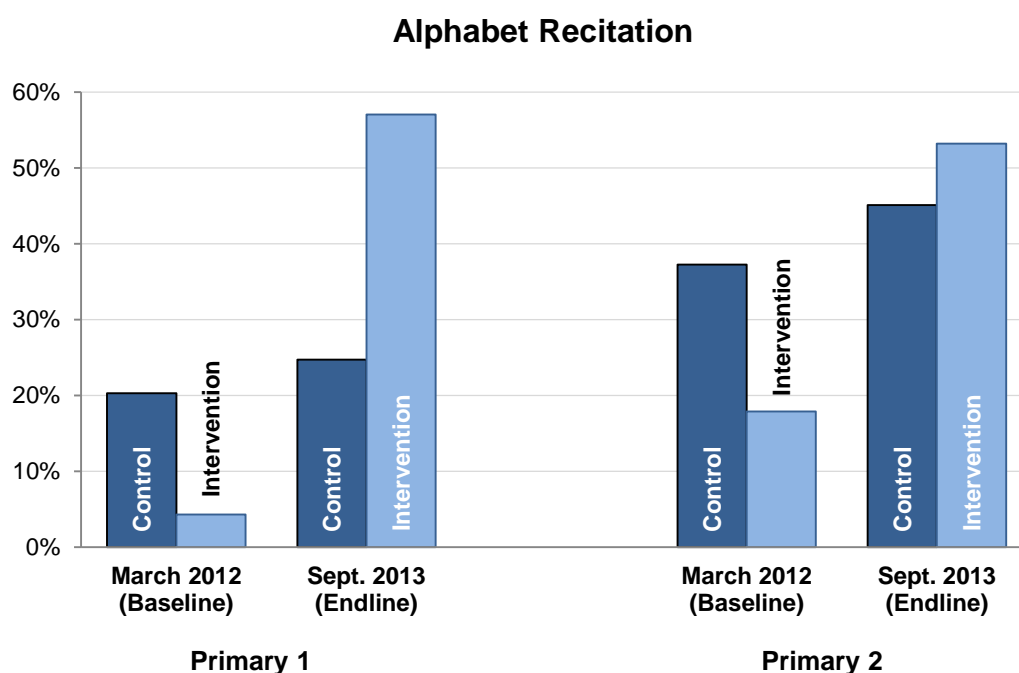
Grouped distributions of correct answers on this subtest showed while both control and intervention P2 students showed a decline in a proportion of students with zero scores, the intervention group's reduction was much more significant: from nearly 80 percent at the baseline to just over 10 percent at the endline. For the control group, the reduction was from 54 percent to just over 40 percent. About 45 percent of intervention students scored between 80 and 100 percent correct on this subtest at the endline, compared to just over 10 percent of the control group students.

Figure 30. Distribution of Percent of Alphabet Recited Correctly by P2 Students

P2 Baseline

P2 Endline
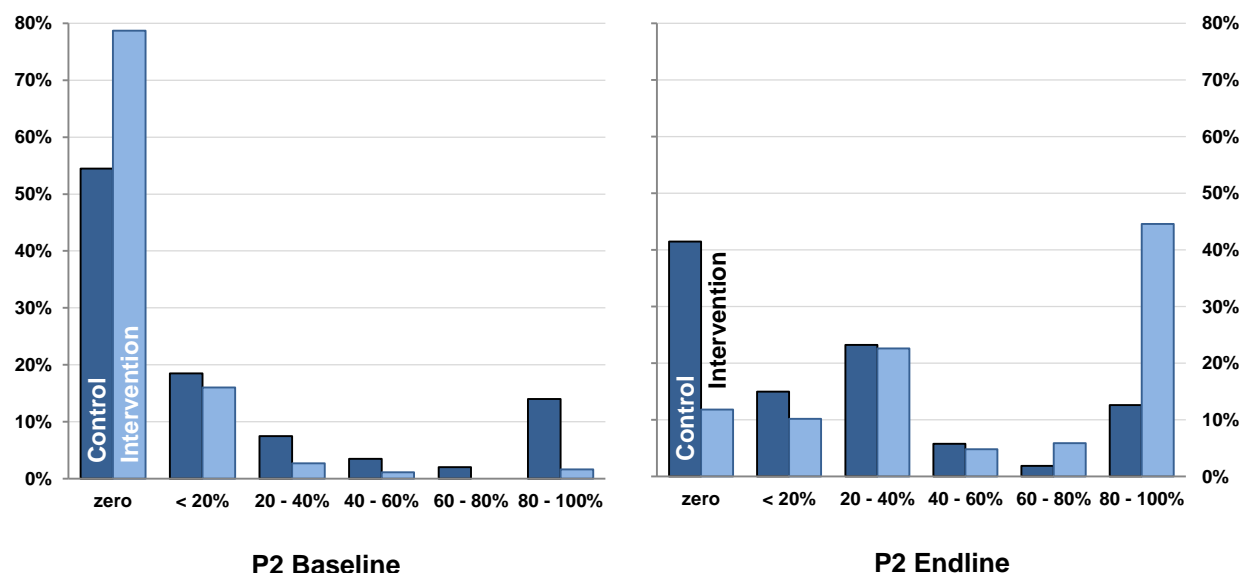
Upper and Lower Case Letter Reading. Students were also tested on letter reading, both in upper case and in lower case letters. Letters in the subtest were presented out of order. Although intervention group students made a significant progress between the baseline and the endline assessments, the overall results remained low, with students of both grades reading correctly fewer than 2 letters, on average. Grouped distributions show that the vast majority of students had zero scores on the reading lower case letters, both at the baseline and the endline. However, intervention group P2 students showed a sizable improvement at the endline.
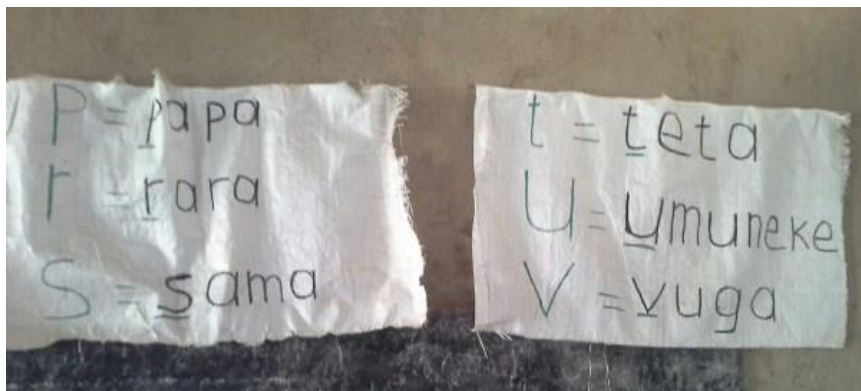
Figure 31. Distribution of Percent of Upper and Lower Case Letters Read Correctly by P2 Students at the Endline

Upper Case Alphabet Reading

Lower Case Alphabet Reading

P2 Endline
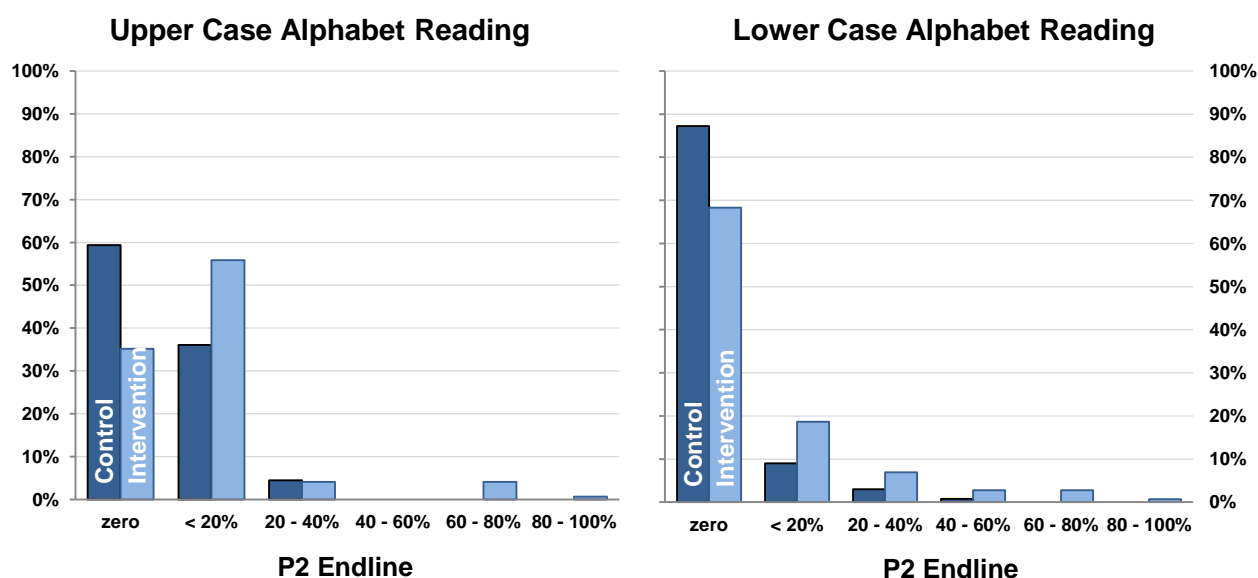
P2 Endline

Finally, the students were asked to read a mix of upper and lower case letters. The test was timed at 60 seconds. Although intervention group showed some gains at the endline, they are really small. Over 90 percent of students still had zero scores at the endline.

Phonemic Awareness. On the phonemic awareness subtest students were asked to identify the first letters in ten words that were read to them, one by one. For example, "what's the initial sound in the word "man"? Students were supposed to answer "mmmm" sound. The graph below showed that there was little difference in achievement between P1 and P2 students on this test. Overall scores deteriorated for all of the P1 students, and for P2 control group students. Intervention P2 students did slightly better at the endline, but not by much. Overall, students answered correctly about 3 out of 10 questions.

Teacher-Made Phonemic Awareness Materials

P = papa
r = rara
S = sama

t = teta
U = Umuneke
V = Vuga

Figure 32. Mean Percent of Correct Phonemic Awareness, by Grade

Phonemic Awareness
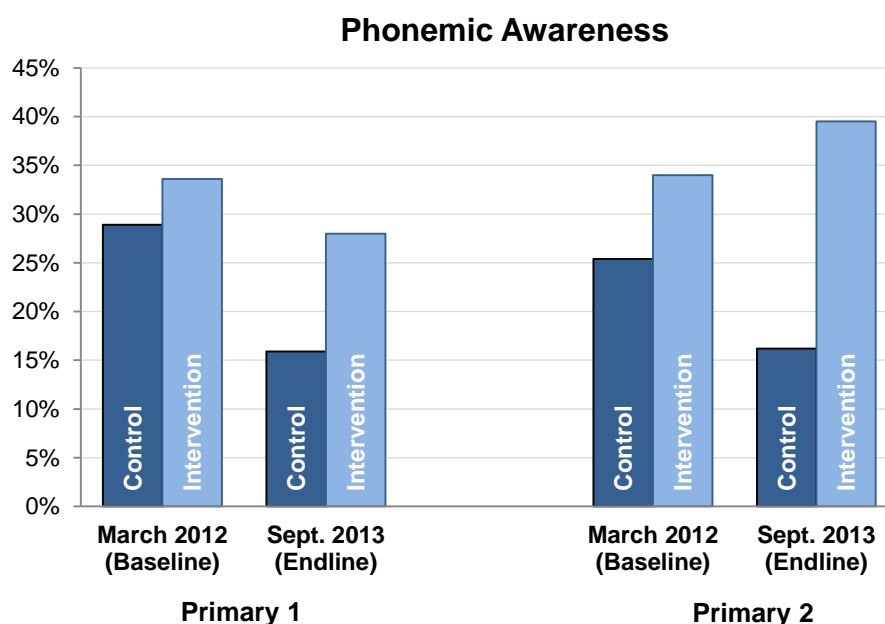
The analysis of the distributions showed that a very high proportion of students had zero scores on this subtest, and a higher proportion of students had zero scores at the endline. Three-quarters of P2 control

students, and 6 in 10 intervention group students failed to answer a single question correctly at the endline on this subtest.

Fluency and Comprehension

literacy assessment included three subtests designed to test students' ability to read and understand words/text:

- Reading familiar words
- Listening comprehension
- Reading a connected text passage

Familiar Word Reading. Familiar word reading subtest measures speed and accuracy of reading commonly used English words. Students may know some or all of the test words by sight, or be able to decode them. All words in the familiar word section are meaningful and frequently used.

At the baseline, neither P1 nor P2 students were able to read the words from the list, within the allocated time of 90 seconds. Both intervention and control group students made some gains between the baseline and the endline, but intervention students gained significantly more, as the graph below demonstrates. The overall level of reading familiar words remained quite low, with 90 percent of the control group P2 students and 70 percent of the intervention group students failing to read a single word from the list. Both boys and girls fared equally poorly on this subtest.

**Figure 33. Distribution of Percent of Familiar Words Read Correctly by P2 Students**

P2 Baseline   P2 Endline
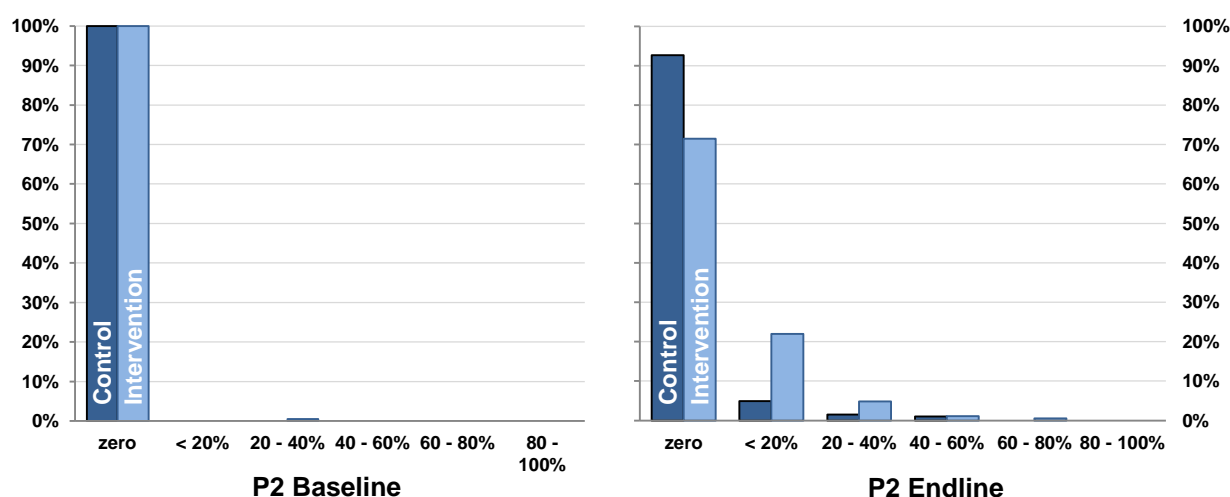
Listening Comprehension. On the listening comprehension subtest, students read a little book called "Mugabo" twice, and asked five comprehension questions. The questions were all content-related and asked about specific facts mentioned in the book.

Both groups had an increase in mean scores between the baseline and the endline, and the intervention group showed a larger increase, compared to the control group. However, the overall comprehension

39

scores are very low. The grouped distributions show that the vast majority of students had zero scores on this subtest, with only a very slight increase at the endline. No differences between genders were observed.

Figure 34. Distribution of Percent of Correct Listening Comprehension by P2 Students


P2 Baseline

P2 Endline
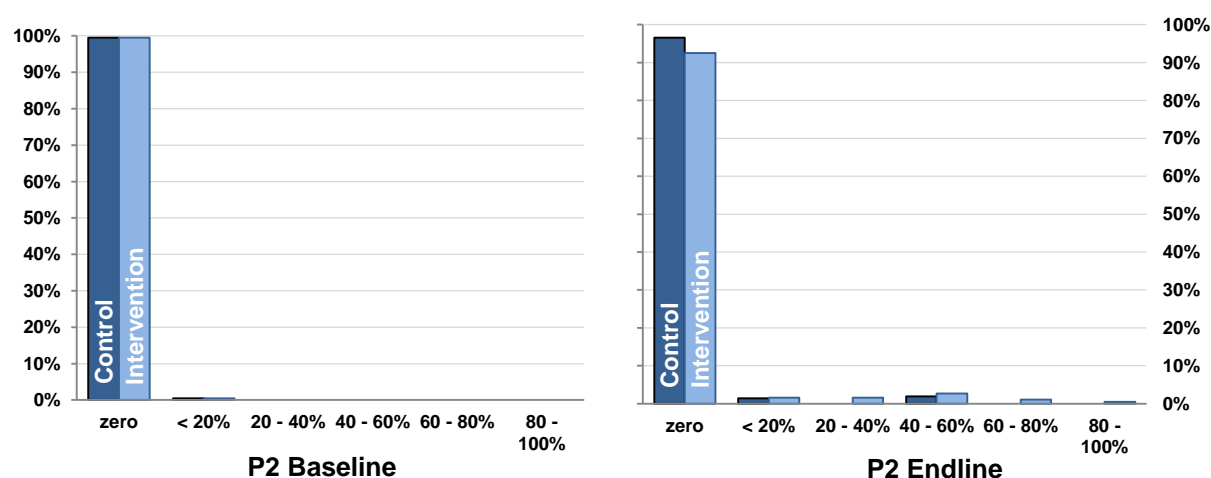
Oral Passage Reading. On the passage reading subtest, students were asked to read a short text that had 29 words. The score included the number of words they read correctly in the passage. Students had a time limit of 60 seconds to read the text. The results presented in the graph below show that very few P1 students were able to read even a single word in the text. Although the endline P1 results are better than the baseline, the mean percent of words read correctly is still very low: just over 1 percent for the control group, and about 3 percent for the intervention group. P2 students did better on this test, and also improved significantly more between the baseline and the endline. However, the mean results at the endline were still low, with just about 11 percent of words read correctly by intervention group students at the endline. Nearly all students had zero scores at the baseline, and 95 and 85 percent of control and intervention group students had zero scores at the endline, respectively.

L3-Provided MP3 Player with Speakers

Students' fluency results were likewise very low. At the endline, control P2 students achieved 1.6 correct words per minute, and intervention P2 students achieved 4.2 correct words per minute, on average. The vast majority of students had zero scores on this measure.

Figure 35. Mean Percent of Words in the Oral Passages Read Correctly, by Grade
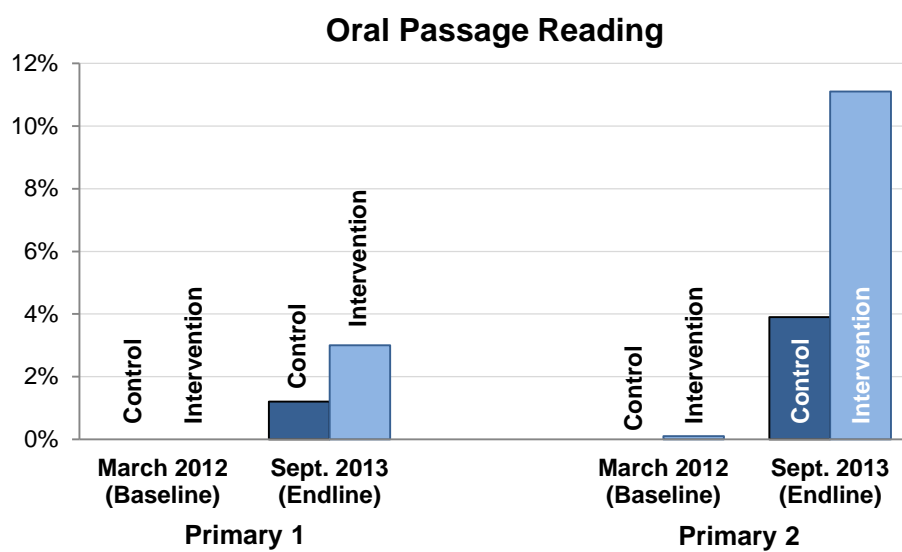

Summary of Findings

P1 students in both intervention and control study groups demonstrated very low skill level in all English literacy assessment subtests but one: alphabet recitation. This is the only subtest on which one of the study groups – the intervention group – scored over 50 percent correct at the endline. The only other subtest on which students scored over 30 percent correct was phonemic awareness. On all subtests a high proportion of students had zero scores. Similarly, P2 students scored on average over 50 percent correct on the alphabet recitation subtest. Two other subtests with scores over 20 percent were basic conversation skills, and phonemic awareness.

The results of the assessment showed a very high proportion of students with zero scores in individual subtests. Although intervention group students did show significant gains at the posttest, the overall scores remained very low.

L3-Provided English Teacher's Guide

Overall, the results of the study showed that L3 intervention is effective for both P1 and P2 students in improving their reading skills in English language.
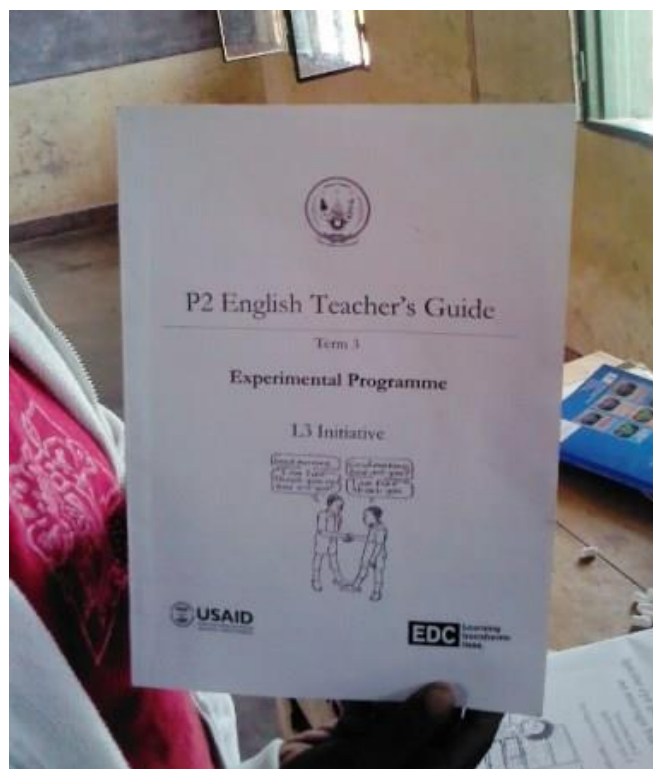
MATH ASSESSMENT FINDINGS

The data analysis of EGMA test results showed that at the baseline control and intervention school students performed at a similar level. However, at the endline intervention school students showed better numeracy skills (at $p < .001$ level for P1 students and $p < .01$ level for P2 students) compared to the control group students. Annex 4 shows details of student achievement on each subtask, by grade, including a percent of learners with zero scores.

The graph below shows the distribution of total mean scores at the baseline and the endline for both grades. The total mean score was computed by adding up the percent correct score for each of the EGMA tasks and then dividing them by the number of tasks. The tasks were not weighted.

Figure 36. Overall Scores Math, by Grade
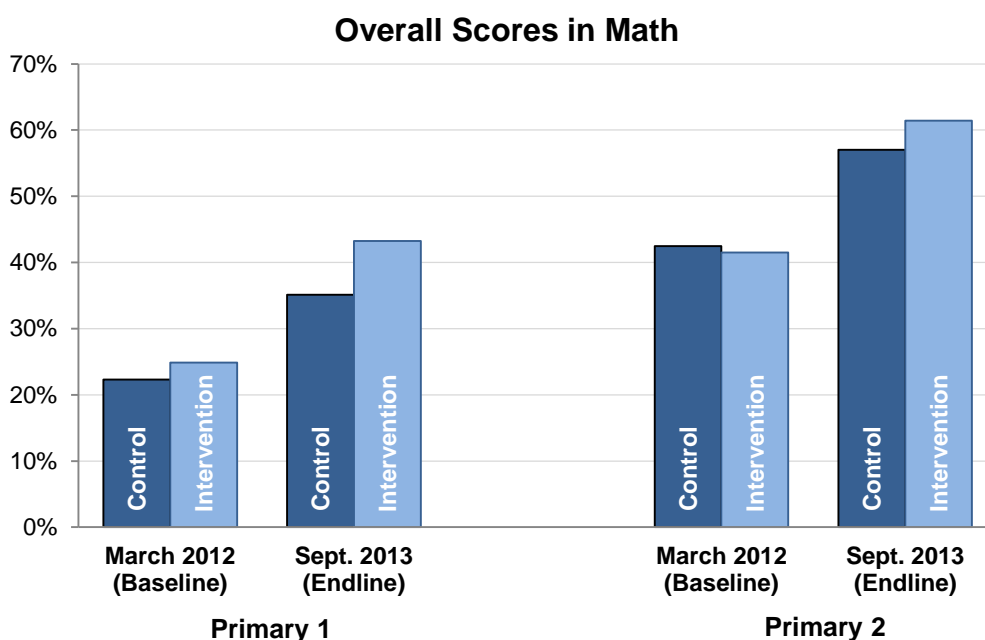
Overall, EGMA test results show that P1 students did well in basic counting, number pairs, number identification, and basic addition and subtraction tasks using manipulatives. P2 students also did well on all tasks, except counting backwards and counting by 2s and by 5s. Students did not do well with tasks requiring operating numbers, such as addition and subtraction without manipulatives.

Annex 5 shows the mean distributions for all subtests, as well as the percent of students with zero scores on each task, and mean scores for students with non-zero scores.

The figure below shows the distribution of the average scores at the baseline and the endline for P1 students, for each of the math assessment subtests. The figure shows the average percent correct achieved by P1 students at the baseline (blue shading) and the gain made at the endline (pink shading).

Figure 37.Summary of Performance on Math Assessment Subtests, P1
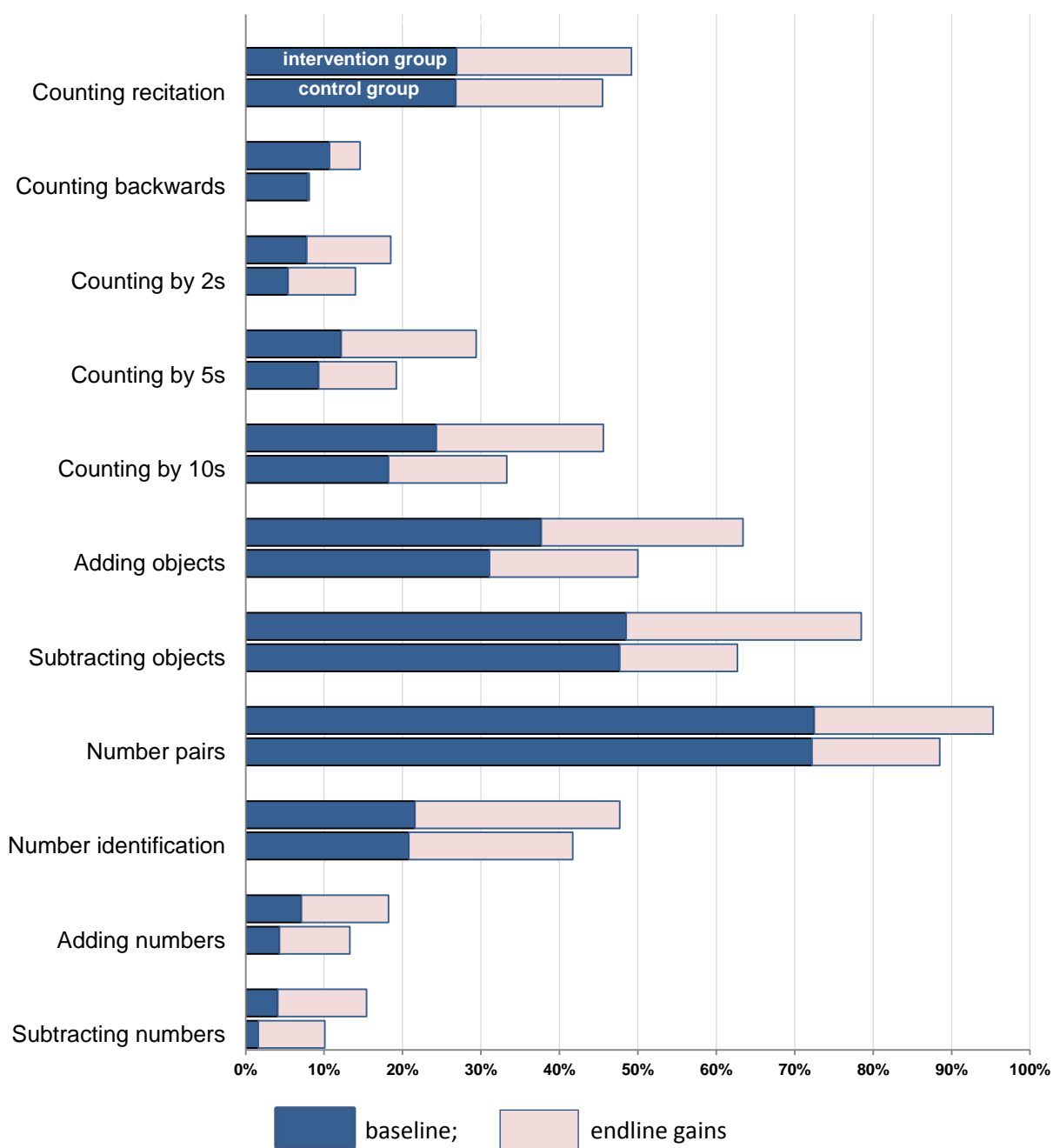
The comparison of means showed that the two study groups in P1 were not equivalent at the baseline, with the intervention group students scoring higher on five out of eleven subtests. At the endline, the intervention group did better than the comparison group on all subtests.

P2 study groups were equivalent at the baseline, with very similar average scores on each subtest. However, at the endline the intervention group scored statistically significantly higher than the control group on four subtests.

Figure 38.Summary of Performance on Math Assessment Subtests, P2
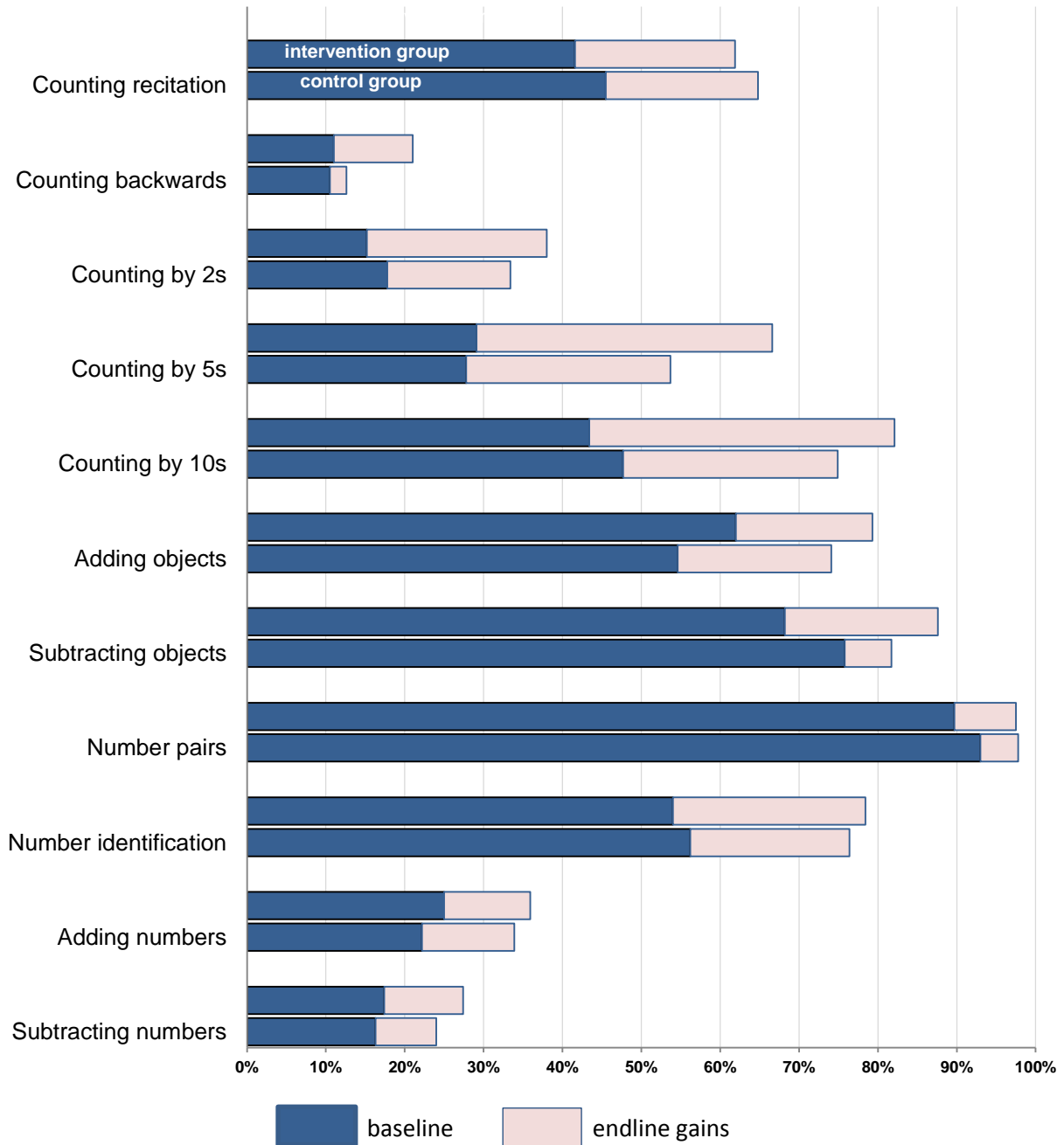
The table below shows a summary of the statistical tests of significance of difference in scores (independent group *t*-test) between the L3 students and the control group students at the endline. As evident from the table, the intervention appears to be particularly effective for P1 students.

Table 6. Summary results of the comparison of means between L3 and control group at the endline

		P1		P2	
		L3	**control**	**L3**	**control**
Foundational math skills	Counting recitation				
	Counting backwards	sig. ($p<.001$)		sig. ($p<.001$)	
	Counting by 2s	sig. ($p<.01$)			
	Counting by 5s	sig. ($p<.001$)		sig. ($p<.05$)	
	Counting by 10s	sig. ($p<.001$)		sig. ($p<.05$)	
	Adding objects	sig. ($p<.001$)			
	Subtracting objects	sig. ($p<.001$)		sig. ($p<.05$)	
	Number pairs	sig. ($p<.001$)			
	Number identification	sig. ($p<.05$)			
Number operations	Adding numbers	sig. ($p<.01$)			
	Subtracting numbers	sig. ($p<.01$)			

The following sections present assessment results by subtest.

Foundation Math Skills

Numeracy assessment included nine subtests designed to measure foundation math skills:

- Counting upwards (zero to 100) , within 60 seconds
- Counting backwards (from the highest reached number in the previous task all the way down) , within 60 seconds
- Counting by 2s (e.g., 2, 4, 6), within 60 seconds
- Counting by 5s (e.g., 5, 10, 15, 20), within 60 seconds
- Counting by 10s (e.g., 10, 20, 30) , within 60 seconds
- Adding objects
- Subtracting objects
- Number pairs
- Number identification

Although some of these subtests were timed, the results presented below are reported as percent correct, not correct per minute.

Counting Upwards. The first to these tasks asked students to count as high as they could go, within the allocated amount of 60 seconds. At the baseline, P1 students counted on average to about 28, and P2 students counted on average to 43. Both control and intervention group students showed an improvement at the endline: P1 students were able to count to about 48, on average, and P2 students counted to over 60.

There is no statistically significant difference between the two study groups either at the baseline, or at the endline.

Figure 39. Mean Number of Counting Upwards Correctly, by Grade

Counting - Recitation
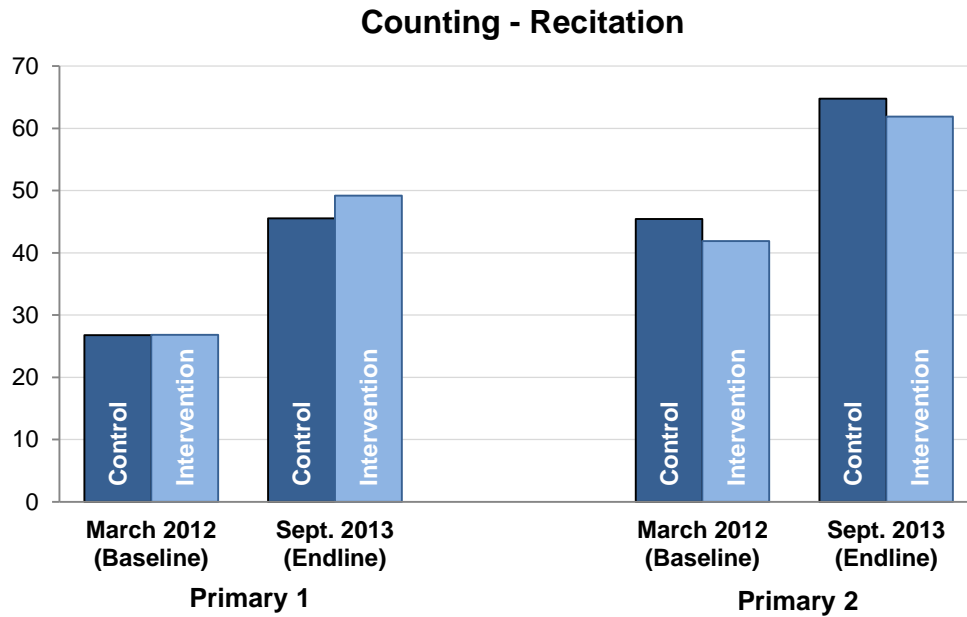
Data analysis showed overall normal score distributions, skewed toward lower scores at the baseline for P1 students, and toward 100 at the endline for P2 students. There are no substantial differences between control and intervention group students on this subtest, or between genders. Very few students failed to count at all.

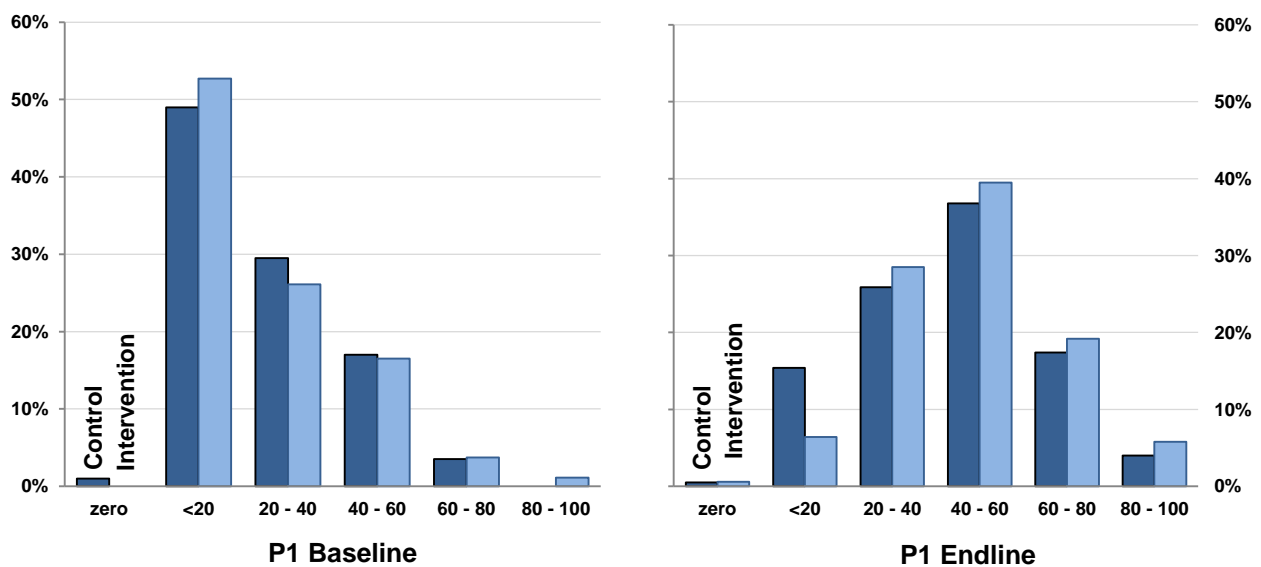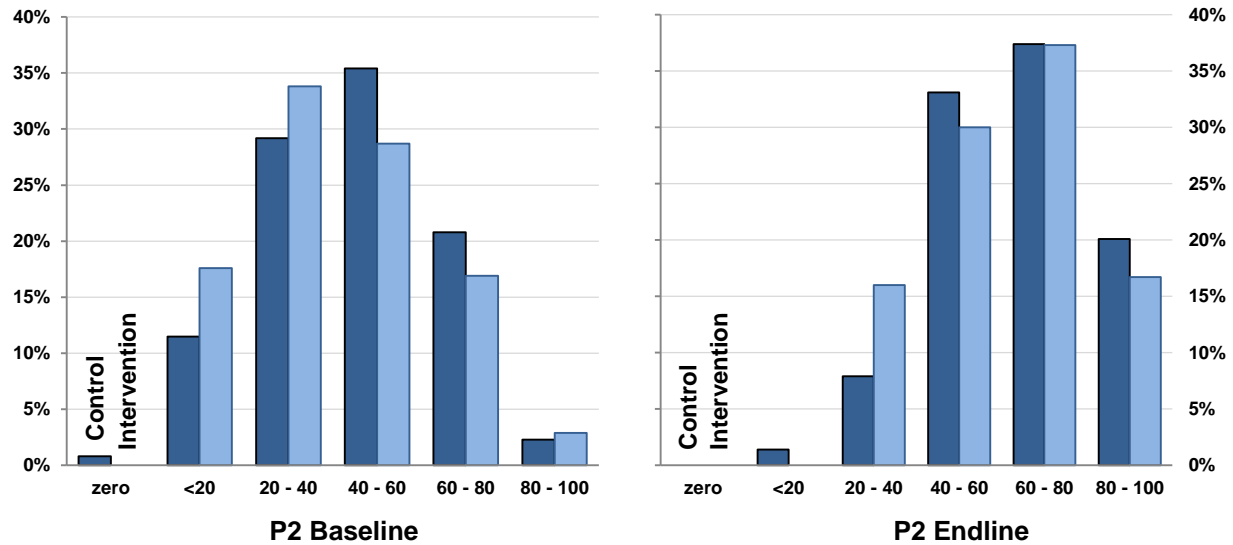**Figure 40. Distribution of Counting Upwards Correctly by P1 Students**

Figure 41. Distribution of Counting Upwards Correctly by P2 Students

Counting Backwards. An ability to count backwards is an important skill that signifies a better familiarity with numbers than simply counting upwards, which students sometimes mechanically memorize. In the test, students were asked to count backwards, starting with the highest number they reached in counting upwards in the previous task.

Results for this subtest were rather low for both grades, although intervention P2 students showed the most improvement between the baseline and the endline. A comparison of means statistical test (independent groups *t*-test) showed that the two study groups for both grades were similar at the baseline, but at the endline the intervention group in both grades scored significantly higher than the control group (at *p*<.001 level).

Figure 42. Mean Percent of Counting by Backwards Correctly, by Grade

Counting by Backwards
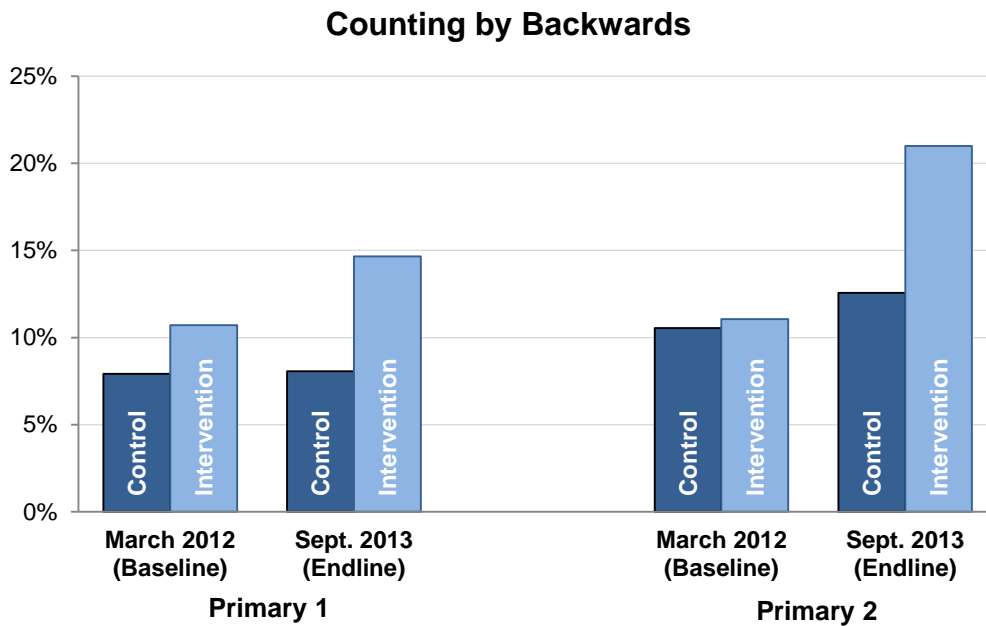
Disaggregation of the results showed that a large proportion of learnings in both grades were able to count less than 20 percent of the numbers that they were supposed to count backwards. A substantial proportion of P1 students were not able to count backwards at all.

Counting by 2s. At the baseline, few P1 and P2 students seem to have been familiar with counting patterns, such as by 2s, 5s or 10s. As a foundational math skill, patterned counting is a very important indicator of students' number sense that is a pre-requisite for successful number operations later on.

Data analysis of counting by 2s showed that very few P1 students had this skill. P2 students did better, particularly intervention P2 students at the endline, who managed to complete just under 40 percent of the task within the allocated amount of time. A comparison of means statistical test (independent groups *t*-test) showed that the two P1 study groups scored statistically significantly higher both at the baseline and at the endline (at *p*<.05 level). No statistically significant differences between the two P2 study groups were found, although the intervention group's gain score was larger than the gain score of control group.

Figure 43. Mean Percent of Counting by 2s Correctly, by Grade
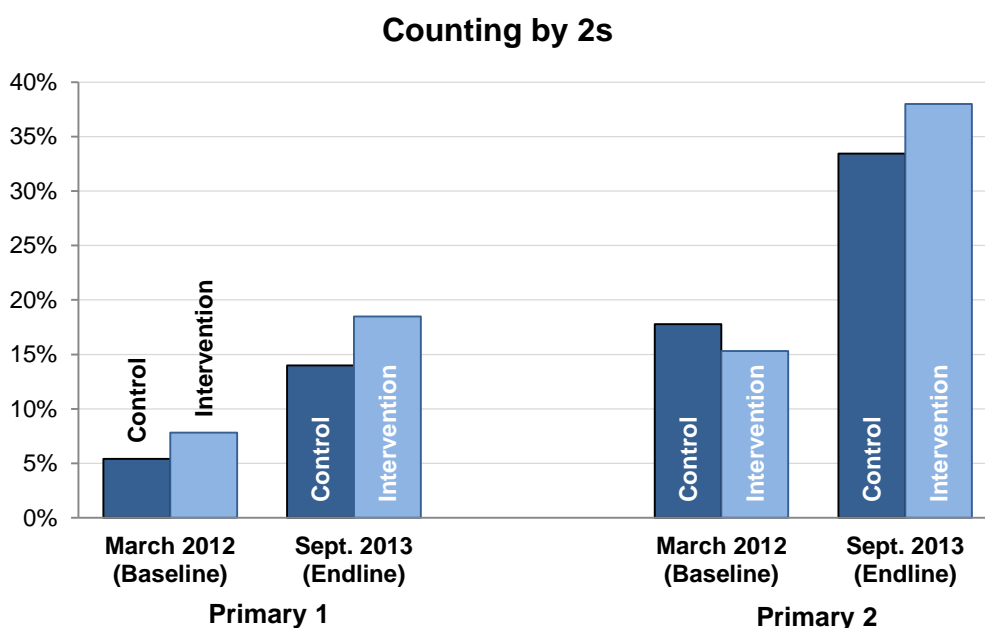
An analysis of score distributions revealed that the majority of P1 students are only beginning to master this skill. At the baseline, nearly all P1 students scored close to zero. The scores improved but slightly at the endline.

At the baseline the majority of P2 students had zero or just above zero scores on this task. However, at the endline the percent of learners scoring below 20 percent went down from three-quarters to just over a third for the control group students, and from 78 to 29 percent for the intervention group students. The distribution at the endline was close to normal, indicating that most student have a limited mastery of the skill.

Figure 44. Distribution of Percent of Counting by 2s Correctly by P2 Students

P2 Baseline
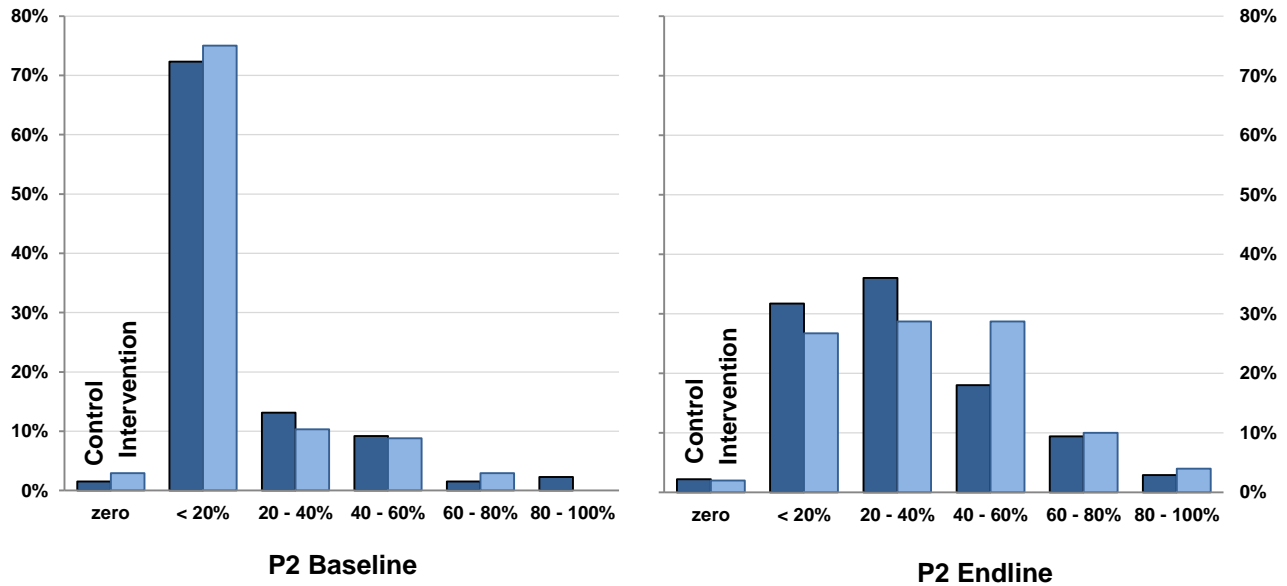
P2 Endline

Counting by 5s. Overall, students seem to be finding it easier to count by 5s than by 2s, since mean scores for counting by 5s subtest are a little higher than for counting by 2s, both at the baseline and at the endline. Students also learned counting by 5 between the baseline and the endline, nearly doubling their original scores. A comparison of means statistical test (independent groups t-test) showed that at the baseline the two study groups were equivalent, while at the endline the intervention group scores significantly higher, at $p<.001$ level, for both grades.

Figure 45. Mean Percent of Counting by 5s Correctly, by Grade
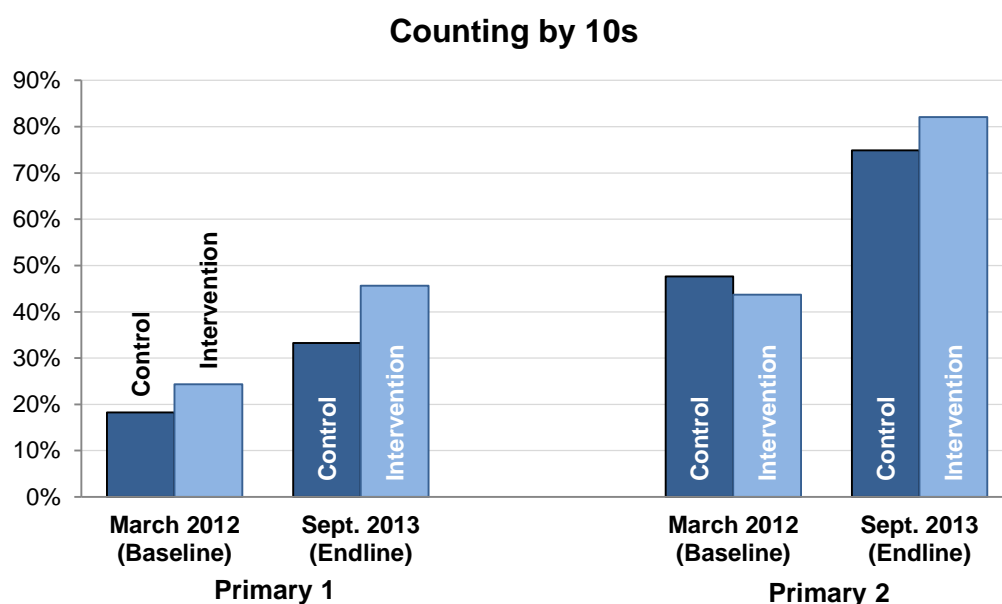

Counting by 5s
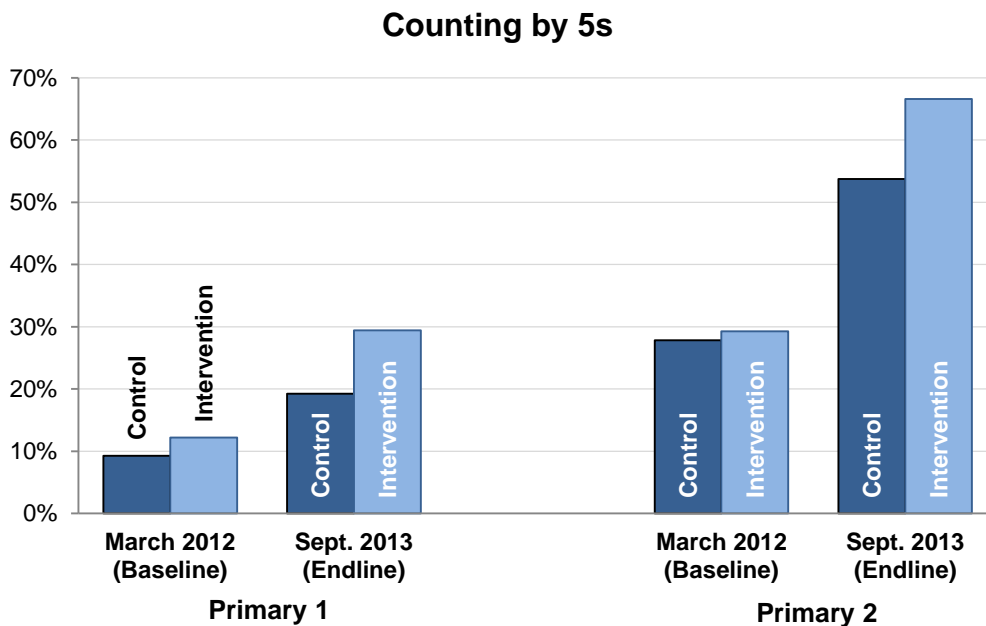
An examination of score distributions showed a skewed toward zero scores distribution for P1 students, and a significantly skewed toward 100 percent scores distribution for P2 students. However, about a third of P2 students scored close to zero at the endline, indicating the need for remediation for these students on this skill.

No substantial differences in performance between boys and girls were found.

Counting by 10s. Of all the patterned counting students did best on the counting by 10s subtest. As the graph depicting the distribution of means across different study groups shows, both P1and P2 students improved between the baseline and the endline. A comparison of means at the endline showed a significantly better performance by intervention group students.

Figure 46. Mean Percent of Counting by 10s Correctly, by Grade

An examination of score distributions showed a normal distribution for P1 students at the endline, with quite a few students reaching 200 (100 percent) within the allocated amount of time. Over half of P2 students reached 200 in counting by 10s within 60 seconds. Very few P2 students did poorly on this subtest. No substantial differences in performance between boys and girls were found.

Operating Numbers with Manipulatives

Young children learn number manipulations using objects, such as toys or fingers, before they can manipulate actual numbers. The tasks were not timed. The test included three tasks using manipulatives:

- Adding stones (e.g, "Here are 3 stones. How many stones do you need to add to your pile to have 5 stones?") – five tasks

- Subtracting stones (e.g., ""Here are 5 stones. How many stones do you need to take away from your pile to have 3 stones?") – five tasks
- Number pairs (e.g., holding 3 fingers up on one hand, "How many fingers are up? How many fingers are down, and not up?") – ten tasks

Data analysis showed that both P1 and P2 students did very well on the subtest with manipulatives, particularly with number pairs subtest. Nearly all students solved all the tasks correctly on this subtest.

To assess students' ability to count using objects, students were asked simple addition questions, using stones. P1 students almost doubled their scores on this subtest and were able to solve between 50 and 60 percent of tasks at the endline. P2 students showed about 50 percent growth between the baseline and the endline and were able to solve about 80 percent of tasks in this subtest at the endline. Students did even better on the subtraction subtest involving manipulatives. Both study groups showed growth between the baseline and the endline, although intervention group improved significantly more than the control group students. The comparison of means test showed that the difference between the two study groups at the endline was statistically significant for both subtests and both grades.

Figure 47. Mean Percent of Adding and Subtracting Objects Correctly, by Grade
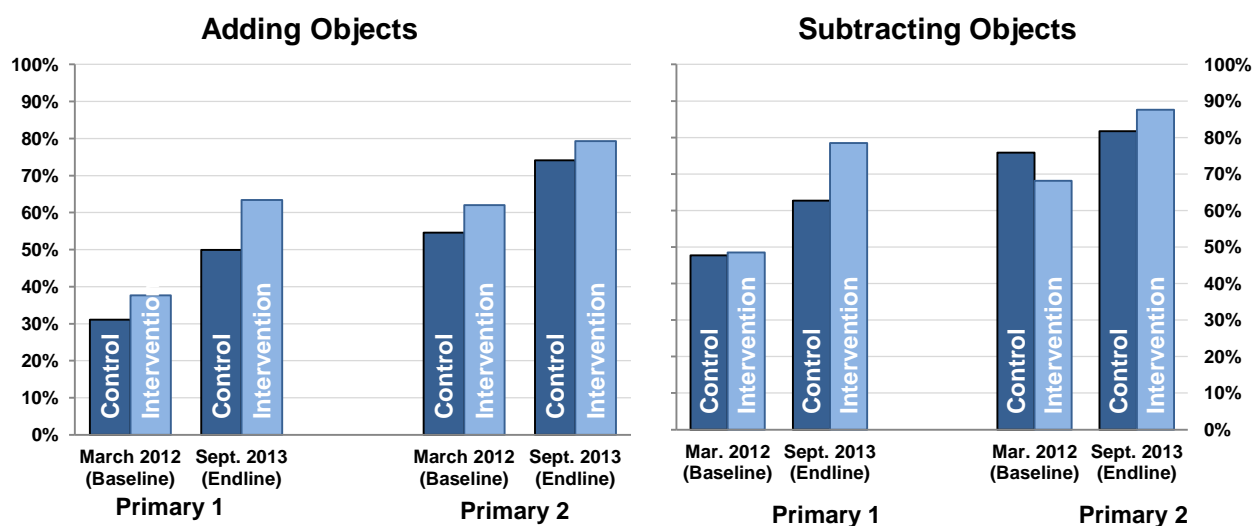
Data analysis of the adding objects subtest for P1 showed the overall change from the distribution with a high proportion of zero scores (about a third of test takers) at the baseline toward a more normally distributed scores at the endline, with fewer than 15 percent students with zero scores, and over a quarter of intervention group students with 80 to 100 percent correct on this subtest. The majority of students fell in the middle of the distribution at the endline.

Figure 48. Distribution of Percent of Adding Objects Correctly by P1 Students

P1 Baseline
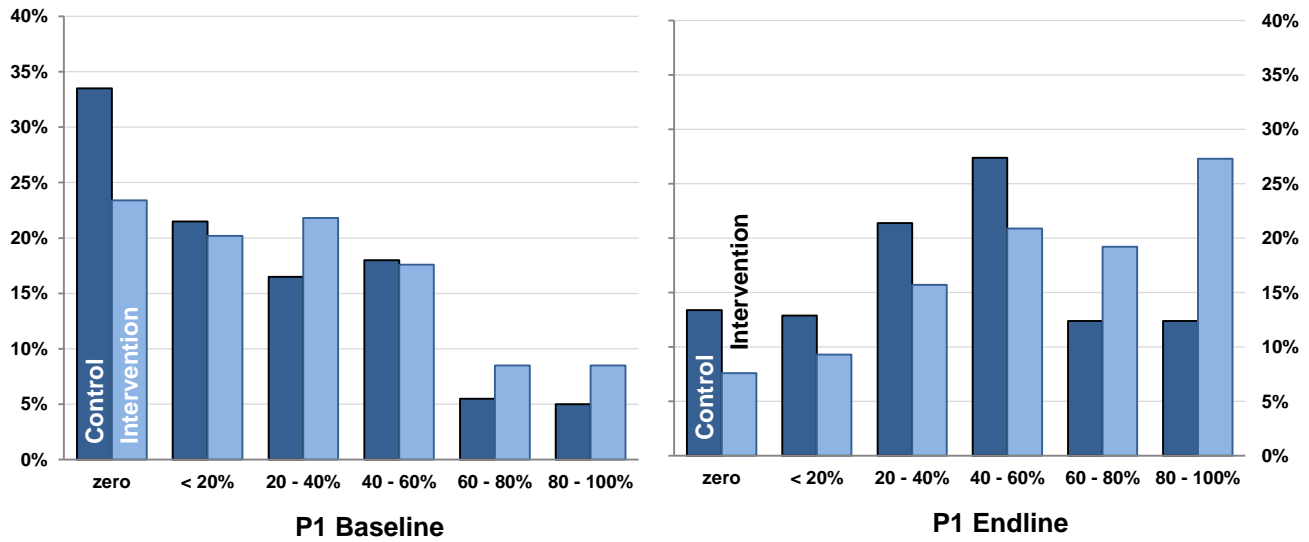
P1 Endline

P2 students from both study groups showed a steady movement toward all correct on this subtest at the endline, with only a few students needing additional support to catch up on this skill.

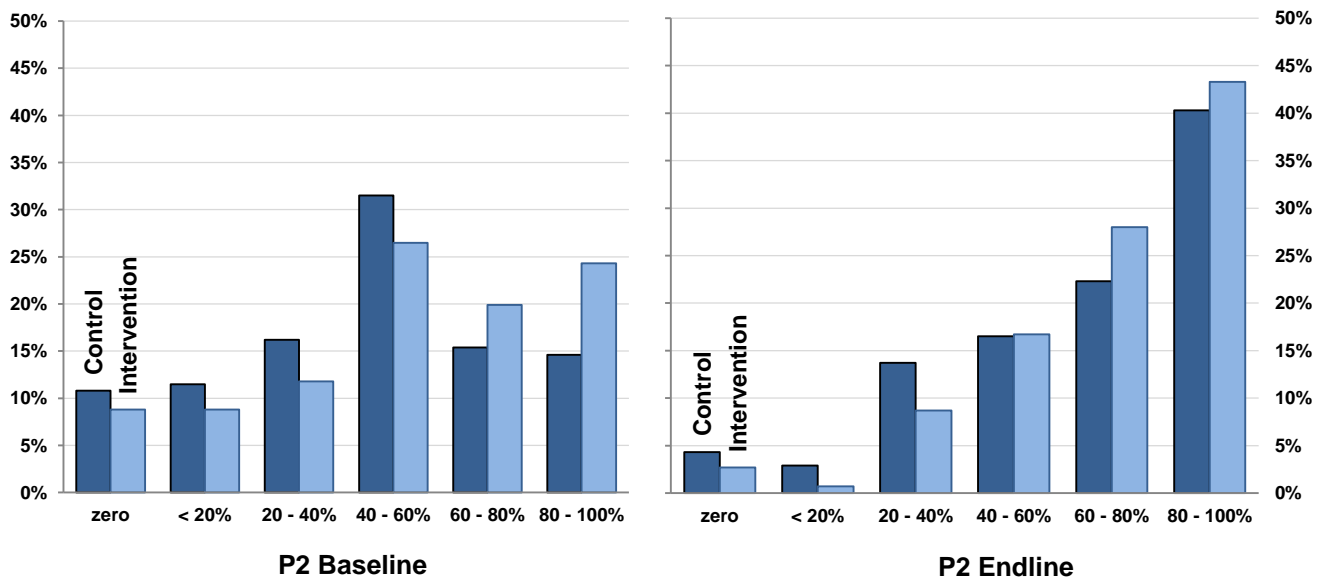**Figure 49. Distribution of Percent of Adding Objects Correctly by P2 Students**

P2 Baseline

P2 Endline

Figure 50. Distribution of Percent of Subtracting Objects Correctly by P1 Students

P1 Baseline

P1 Endline
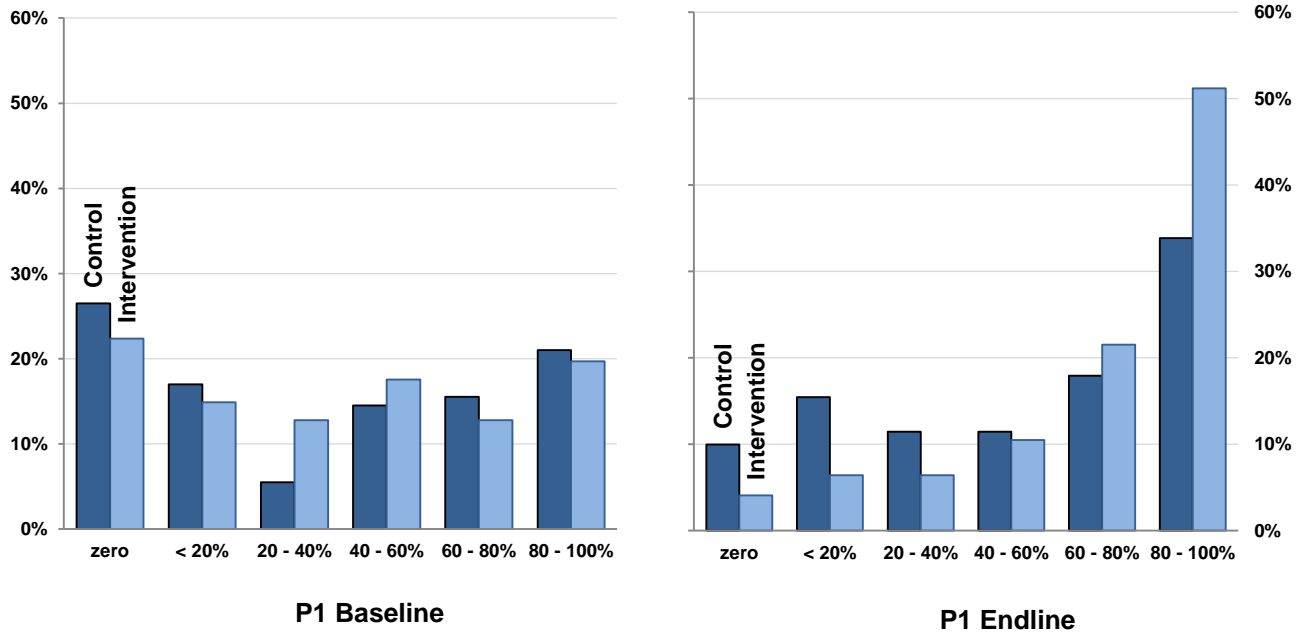
Figure 51. Distribution of Percent of Subtracting Objects Correctly by P2 Students
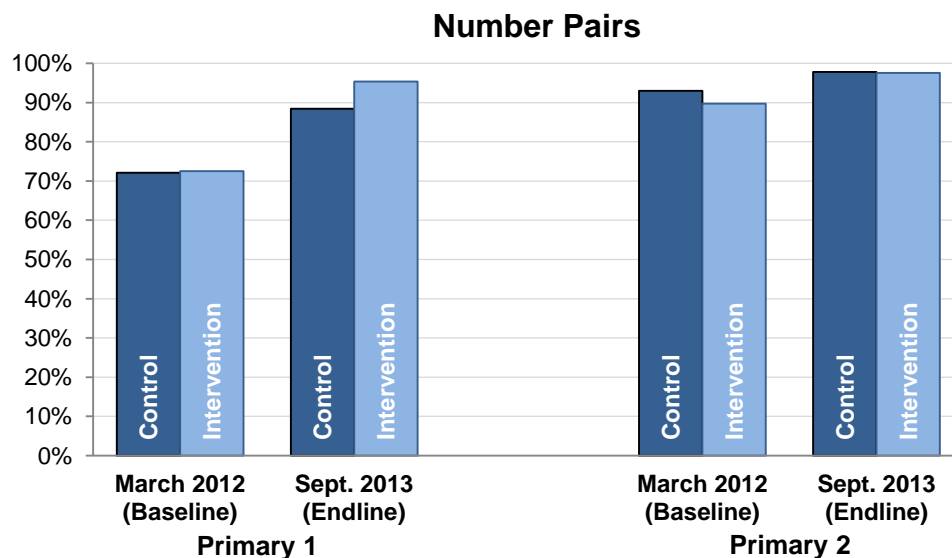

P2 Baseline

**P2 Endline**
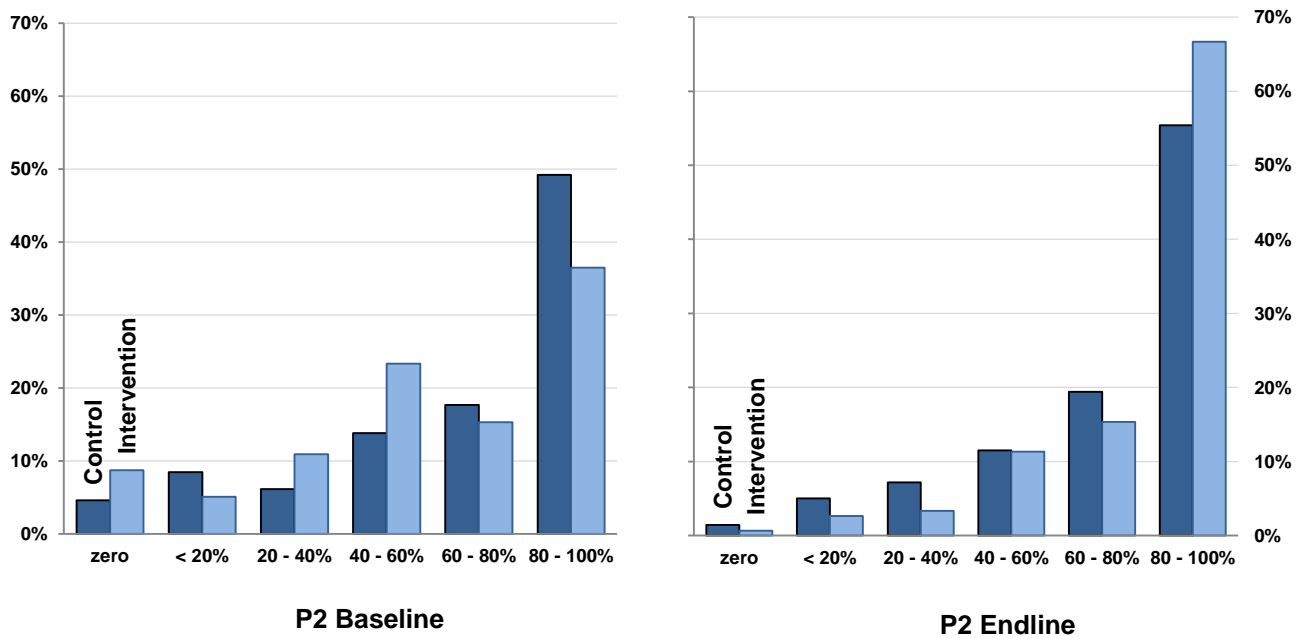
L3-Provided Counting Manipulatives

Number Pairs. Number pairs tasks involved counting how many fingers the assessor held up, versus how many were down. The subtest consisted of five questions asking about a number of fingers up, and 5 questions asking the learner to identify how many fingers were *not* up. The graph below shows the distribution of mean scores for this task across test groups. The results show that both control and intervention P1 students improved between the baseline and endline. Since nearly all P2 students solve problems in this task correctly at the baseline, there was little growth between the baseline and the endline. The difference between the means of the two study groups at the endline was statistically significant.

**Figure 52. Mean Percent of Number Pairs Tasks Solved Correctly, by Grade**

Grouped distributions below show that the vast majority of P1 students were able to solve these tasks at the baseline as we as at the endline. Nearly all P1 students were able to solve the ten number pairs problems so the grouped distributions are not displayed.

Figure 53. Distribution of Percent of Numbers Paired Correctly by P1 Students
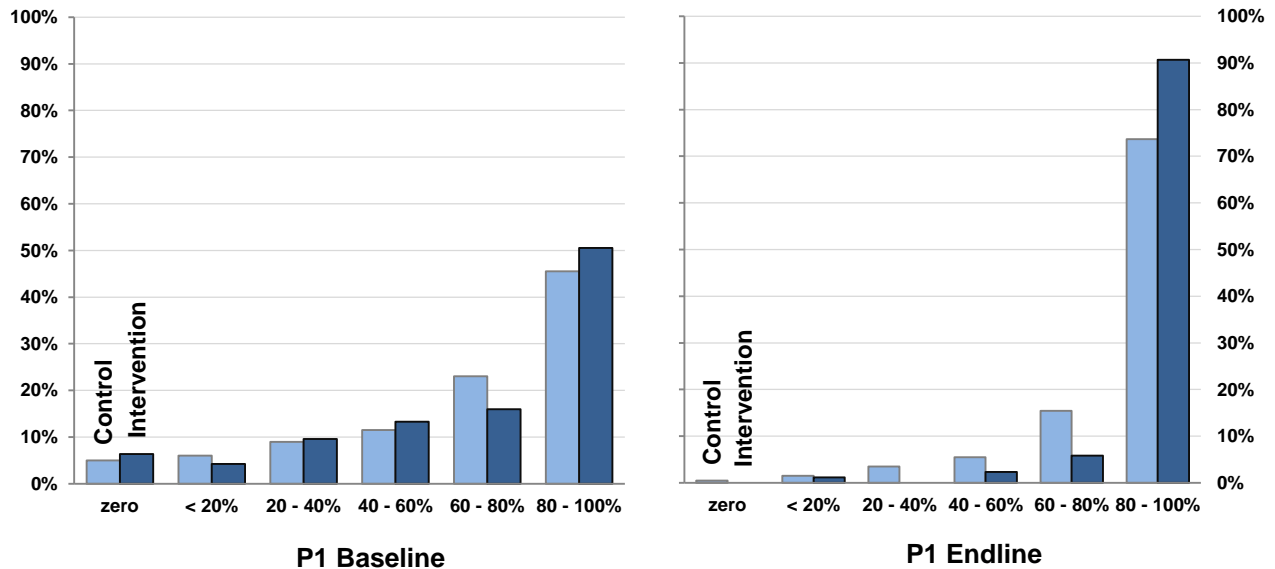
Number Identification. To assess students' familiarity with numbering notation, students were asked to identify numbers written as a list of 100 numbers, within 60 seconds. As evident from the graph below, P1 students were able to name about 20 percent of the numbers at the baseline and about half of them at the endline. P2 students identified about 80 percent of numbers correctly at the endline.

Figure 54. Mean Percent of Numbers Identified Correctly, by Grade

56
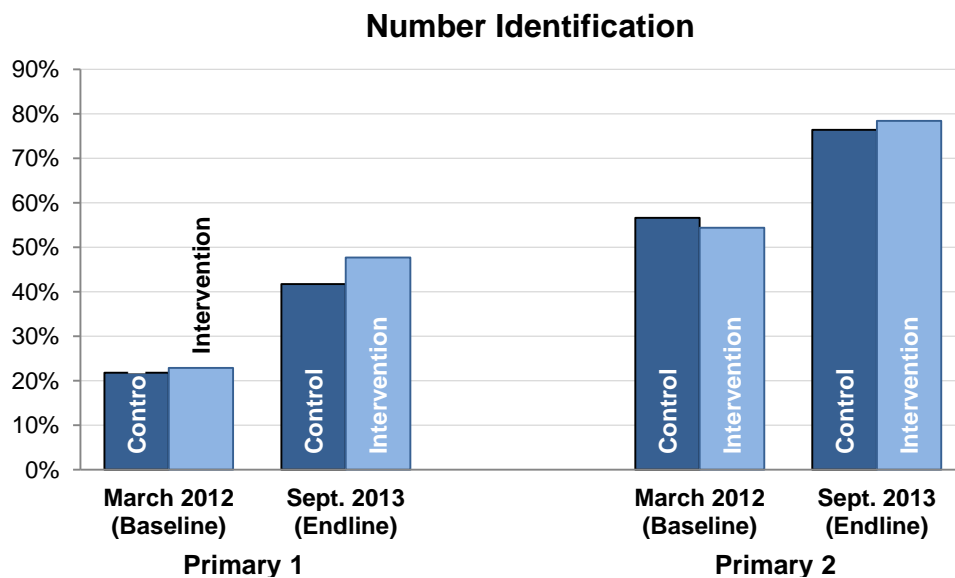
Number Operations: Adding and Subtracting

The math assessment included two subtests designed to measure students' skills in basic single and double digit number operations, such as addition and subtraction. The graphs below show the distribution of mean responses, by study groups. Students did significantly better at addition than subtraction. The intervention group students demonstrated statistically significantly better performance on both subtests.

Figure 55. Mean Percent of Adding and Subtracting Numbers, by Grade
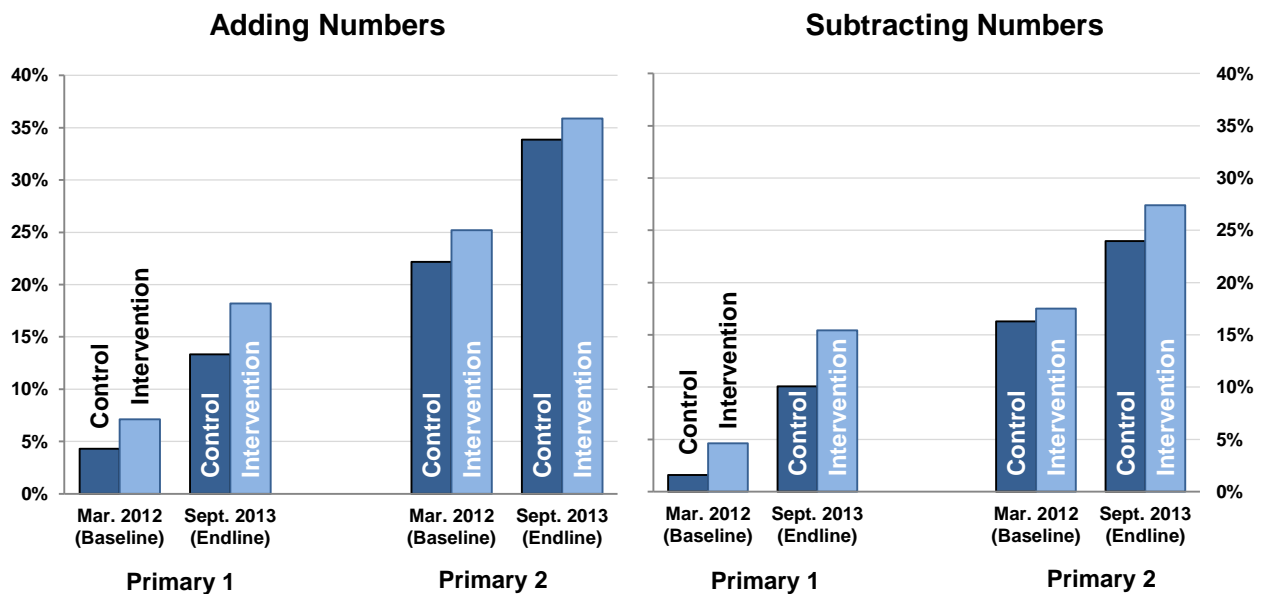
Grouped distributions below show a large proportion of students with zero scores, both among P1 and P2 students. The improvement between the baseline and the endline among P1 students was manifested itself mostly in reduction of the proportion of students with zero scores. Similar trend was observed with P2 students. The distribution of scores of P2 students showed that the majority of students could solve some, but not all 18 addition problems presented to them.

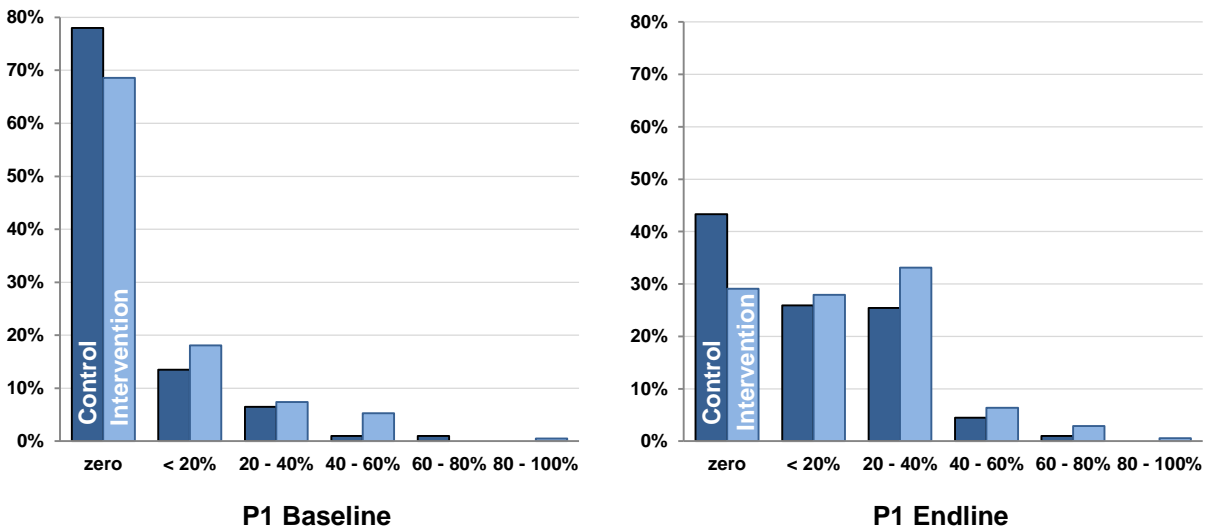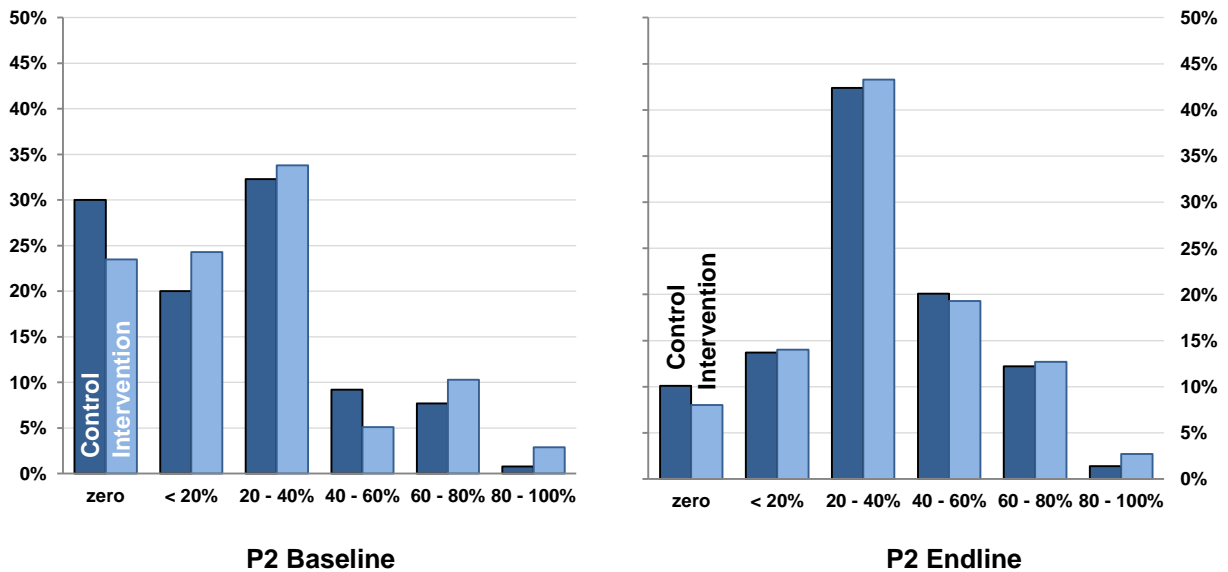**Figure 56. Distribution of Percent of Adding Numbers Correctly by P1 Students**

P1 Baseline

P1 Endline

Figure 57. Distribution of Percent of Adding Numbers Correctly by P2 Students

P2 Baseline

P2 Endline

Close to 90 percent of P1 students at the baseline could not solve a single subtraction problem. However, this number was reduced by about a half at the midline. Quite a few of P2 students were able to solve some of the 18 subtraction problems on the test.

Figure 58. Distribution of Percent of Subtracting Numbers Correctly by P1 Students
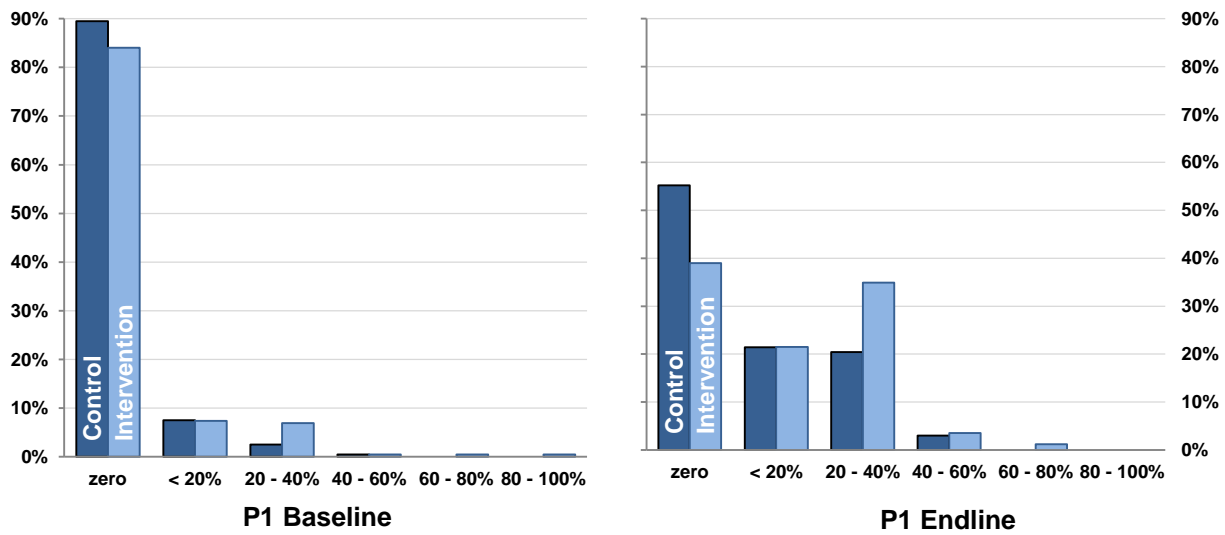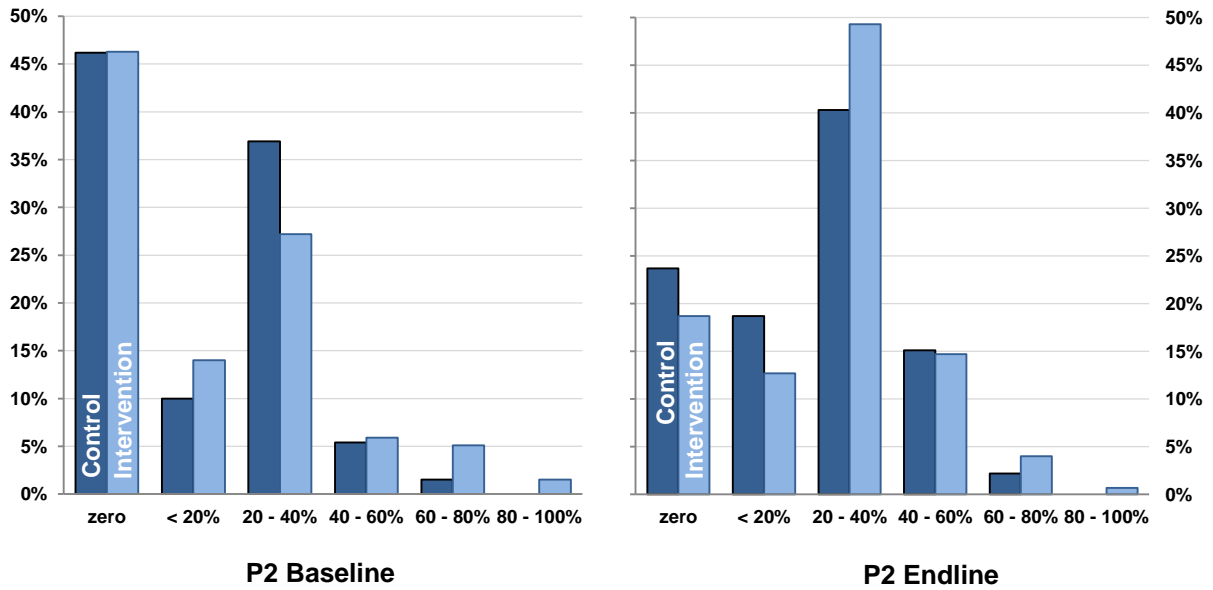

P1 Baseline

P1 Endline

Figure 59. Distribution of Percent of Subtracting Numbers Correctly by P2 Students

P2 Baseline

P2 Endline

Summary of Findings

At the endline, P1 students in both intervention and control study groups scored over 50 percent, on average, in three out of eleven subtests of the assessment:

- Adding objects
- Subtracting objects
- Number pairs

While many P1 students completed individual subtests without any errors, a substantial proportion of students had zero scores. The tasks on which students demonstrated proficiency are less abstract and have relevancy in their lives. They also provide foundational skills for number operations later on, but at the time of the assessment few students demonstrated mastery of operating numbers, not objects. Particularly noteworthy is a significant proportion of students with zero scores on the number identification subtest.

P2 students scored on average over 50 percent correct on the following subtests:

L3-Provided Math Materials

- Counting (recitation)
- Counting by 5s
- Counting by 10s
- Adding objects
- Subtracting objects
- Number pairs
- Number identification.

Similar to the results of the P1 student assessment, a high proportion of students had zero scores in individual subtests. Although some subtests showed close to normal distributions of scores among students with non-zero scores, a high percent of students with zero scores skews overall results. Similar to literacy assessment findings, math assessment results show that Rwandan classrooms are filled with students with very different math skills, however, the contrast between lower and higher achieving students is not as sharp as in literacy.

The overall results of the pilot show that L3 intervention is effective in improving student achievement in math. Similar to literacy assessment results, the intervention was found to be particularly effective at P1 level.